Detecting large-scale structure in the era of petabyte/gigaparsec Astronomy

Murphy, David, Niall, Adrian

Detecting large-scale structure in the era of petabyte/gigaparsec Astronomy

David Niall Adrian Murphy

A thesis submitted to the University of Durham
in accordance with the regulations for
admittance to the Degree of Doctor of Philosophy.

Institute for Computational Cosmology
Department of Physics
University of Durham
April 2011

Detecting large-scale structure in the era of petabyte/gigaparsec Astronomy

David Niall Adrian Murphy

Submitted for the degree of Doctor of Philosophy

April 2011

Abstract

In this thesis, we present a study of the identification of large-scale structure in optical astronomical surveys. This encompasses the detection of large connected structures of galaxies in spectroscopic datasets and of galaxy clusters in deep photometric surveys.

Beginning with a survey featuring full 3D galaxy data, in chapter 2 we present a method to identify filamentary structure after accounting for the line-of-sight velocity distortions characteristic of the virialised systems we search for. We compare data from a real galaxy survey to a series of realistic mocks. Despite broad similarities between the two, we find models do not reproduce the largest observed structures. To evaluate the exploration of a multi-band survey lacking spectroscopy, we simulate the effects of photometric redshift uncertainties on galaxy redshifts. Our findings provide limits on the accuracy of photometric redshift estimators required to recover the same diverse range of structures detected in the original spectroscopic survey.

As an alternative means of exploiting the deep multi-band photometric data common to wide-area observational campaigns, in chapter 3 we present a red sequence-based algorithm to detect galaxy clusters with Voronoi diagrams. This algorithm makes no prior assumptions about cluster properties other than the similarity in colour of their members, and an enhanced projected surface density. Testing the algorithm with mock galaxy survey data reveals a detection performance equalling or exceeding that of alternative detection algorithms.

Chapter 4 describes the application of this algorithm to a 270deg^2 survey with deep SDSS photometry. The scientific exploitation of 4,000 $\mathrm{z} \leq 0.6$ cluster detections from this survey is ongoing, but work presented here shows: agreement with the red sequence slope evolution derived from semi-analytic galaxy models, evidence stellar age is not responsible for responsible for the sequence slope, and a well-defined colour-colour track of potential use in photometric cluster redshift estimation. We detail improvements made to the cluster algorithm in chapter 5. Through a series of case studies we verify our approach successfully identifies galaxy clusters in a diverse range of surveys, from volumes spanning $2h^{-3}\mathrm{Gpc}^3$ to deep near-IR detections at $\mathrm{z} \sim 1$. Based on our findings, we expect the Pan-STARRS 3π large-area survey to identify over 10^5 clusters and groups.

In chapter 6, we explore the characteristics of randomly-distributed noise in Voronoi diagrams. We verify the model traditionally used to describe the distribution of Voronoi cell areas in Poisson data fails to describe the frequency of high-density random cells. Because high-density cells resemble those expected from a population of galaxy cluster members, using a large dataset generated in this study we propose an alternative model that better estimates the frequency of their areas. This new model may in the future be used to improve Voronoi-based recovery of clustered data in a diverse range of applications, both astronomical and otherwise.

Dedicated to

Daphne Cecilia Williams

Contents

List of Figures

List of Tables

Declaration

The work described in this thesis was undertaken between 2006 and 2010 while the author was a research student under the supervision of Prof. Richard Bower, Dr. Vincent Eke and Prof. Carlos Frenk in the Department of Physics at the University of Durham. This work has not been submitted for any other degree at the University of Durham or any other University.

Portions of Chapter 2 of this thesis have been submitted in the form of a paper:

- **David Murphy**, Vincent Eke, Carlos Frenk , "Filamentary structure in the 2dF-GRS", 2011, MNRAS, 413, 2288

Portions of Chapter 3 of this thesis have been submitted in the form of a paper:

- **David Murphy**, James Geach and Richard Bower, "ORCA: The Overdense Red-sequence Cluster Algorithm", 2011, MNRAS, in press (arXiv/1109.3182)

Portions of Chapter 4 of this thesis have been submitted in the form of a paper:

- James Geach, **David Murphy** and Richard Bower, "4098 galaxy clusters to z$\sim$0.6 in the Sloan Digital Sky Survey equatorial Stripe 82", 2011, MNRAS, 413, 3059

Acknowledgements

The path from student to researcher is unique. My first brush with "big astronomy" came amid the narrow streets and ancient fortifications of Valencia's old quarter, where the cutting edge of research was laid bare to an undergraduate who knew what a galaxy was, but not how to study it. I thank Ofer Lahav for setting me on the path that led here.

I was privileged to have my postgraduate experience at Durham guided and shaped by three very capable supervisors - Richard Bower, Vincent Eke and Carlos Frenk. To these three I owe an immense debt of gratitude; it is only through their tireless dedication, enthusiasm and guidance I have reached this point. I can only hope to go on and inspire future researchers with the same energy, experience and vision. My particular thanks to Richard, for his patience and assistance in the preparation of this work.

I pay particular homage to Lydia Heck - frequently beyond the call of duty, but always with a smile. Your help, humour and insight have played an integral part in the work presented here, and long may it do so for future students lucky enough to study at the Institute. Alan Lotts deserves special mention for donning the safety helmet as I crashed my way through a variety of machines over the past few years.

I am particularly grateful to, among others, John Lucey, Alastair Edge, Nelson Padilla, Russell Smith, Carlton Baugh, Chris Stubbs, Ryan Chornock and Alexie Leatheaud for the kind use of data, the imparting of valuable (and often heeded) advice, inspiration and discourse over coffee. A particular mention goes to Peter Draper and Nigel Metcalfe for their tireless contribution to Pan-STARRS, and in turn to my thesis.

My acknowledgements would be incomplete without a tip of the hat to the many postgrads and postdocs who have enriched my time at Durham. A special mention to Rob Crain, Allen Shone, Alvaro Orsi and Dan John; I will treasure the diversions, jokes, debates, whiteboard miscellany and many coffee breaks that made the office a great place to work. Beyond the intellectual powerhouse of OC320, and in no particular order, my appreciation to Milan Raičević, Nikolaos Fanidakis, Claudia Lagos, Juan Gonzales, Jae-Woo Kim, Tim Rawle, Nikolaos Nikoloudakis, Elise Jennings, Alex Merson, John Helly, Kristen Coppin, Neil Crighton, John Stott, Mark Swinbank, Violeta Gonzalez, Gabriel Altay, Ryan Hickox and anyone else I unintentionally missed. In particular, I must single out Jim Geach for his enthusiasm, guidance, and two particularly memorable Pan-STARRS meetings.

Closer to home, David Eate, almost entirely to blame for getting me interested in computers, has my thanks for making my studies altogether less daunting through his early encouragement of what is now a bad habit.

Jenny - my inspiration, distraction and sometime tormentor! Thank you for your enduring patience, unshaken support and comfort. Especially for being there during the ups and the downs, and for the times I needed you the most.

To my grandmother Margot, provider of much encouragement, love and hints for the weekend crossword, often through the medium of text message. To my wider family - for the times I forgot, for the times I couldn't be there, and for the times I was, but was busy - thank you for your understanding and encouragement.

Finally, my eternal thanks, gratitude and love to my parents Dawn and Adrian. To my Mother, who will tell you in fact my first brush with "big astronomy" came from the milk bottles that space rockets are launched from, and my Father, who still looks for the Mask of Apollo in all I do (hint - page 13). Your dedication, devotion and unwavering belief in me made this possible; this work is the product of your success too. Thank you.

Chapter 1 | *Introduction*

1.1 History of Astronomy

The science of Astronomy is rooted in the observed regularity of the cosmos. Although the Babylonians and Egyptians were highly proficient in recording the apparent motion of objects in the sky, their motivation and understanding was tied to mysticism and religion.

The ancient Greek society was the first to question their conception of the world, and to distinguish between mythical and physical descriptions of the cosmos. Critical thought free from religious dogma allowed convergence in understanding to the centrifocal universe models described by Plato and Aristotle (350BC). In these, the motion of the Earth was transferred to the heavens, requiring an account of both genuine and Earth-based motion. Developed extensively by Ptolemy (120), but originating in the work of Hipparchus and Apollonius (135BC), the eventual adoption of the epicycle accounted for observed retrogression of the planets. Epicycles, complemented later by the eccentric and equant, could account for nearly all of the observable motions of the astronomical bodies.

Some time earlier (280BC) Aristarchus proposed a heliocentric model of the universe. Robustly rejected on the grounds of observational evidence (parallax was at that stage unobservable), it also questioned the ancient (yet secular) dogmas from which the Ptolemaic model originated. As the centrifocal system grew, it encompassed a series of progressively convoluted amendments and additions. Failure to account for phenomena observed by the first generation of telescopes led to a gradual, but inevitable collapse of the model. It was at this point the heliocentric work of Nicolaus Copernicus (1543) began to receive serious consideration.

Galileo Galilei's discoveries (1610) of three bodies orbiting Jupiter and of the phases

of Venus destroyed the Aristotelian notion of the universe. Parallel to this, Johannes Kepler's empirical laws of planetary motion (1609) provided the theoretical framework for the dynamics of the solar system. Isaac Newton would later derive these in *Philosophiæ Naturalis Principia Mathematica* (1687), his work on the law of universal gravitation that successfully unified Astronomy and Physics. Whilst the paradigm shift towards heliocentrism did not entirely resolve the traditional objections to a heliocentric Copernican universe, it ultimately represented the transition to a simpler, more elegant model.

1.2 The renaissance era of Astronomy

Edwin Hubble's landmark observations (1922-1923) on the recessional velocities of distant galaxies sparked a considerable shift in the way the universe was understood, and introduced the notion of an expanding cosmos. At that time, there was significant uncertainty and debate over the distance to these nebulae. Arguing for a Milky Way encompassing the entire universe, in 1920 Harlow Shapley debated Heber Curtis' notion of distinct "island universes", of which the Milky Way was but one. Observing the spectra of spiral "nebulae" thought to be local, Hubble noted spectral features similar to those of stellar objects but shifted to longer wavelengths. This doppler-shifting of the light, observed in all directions, suggested these objects were receding from the Earth. Connecting this observation to the identification of faint Cepheid variables (robust measures of distance known as *standard candles*) in these nebulae revealed they were at distances far beyond the size of the Milky Way. In 1929 Hubble related spectrally-inferred recessional velocities (v) to the luminosity-inferred distances (r) of the "island universes" via his eponymous law:

$$v = H_0 r \tag{1.1}$$

where H_0, Hubble's Constant[1], describes the expansion rate of the universe. The linearity in this relation is well defined out to redshifts of z$\simeq$0.25.

The theoretical framework for an expanding universe was laid down by Friedmann in 1922, but it was developed by Georges Lemaître. Lemaître argued that if the universe was indeed expanding, then it must have been smaller in the past. This implied the

[1] A constant in space, but not time. The 0 subscript denotes H(t) at the present epoch.

universe once existed in a phase he described as the "primeval atom"; a hot, dense state we refer to today as the big bang.

In the 1950s, Ralph Alpher and George Gamow argued the cooling afterglow of the big bang should be observed as a background radiation field corresponding to a temperature of around 5K. Some 300,000 years following the big bang, expansion of the universe cooled the primordial baryonic plasma to temperatures permitting the out-of-equilibrium recombination of free electrons with nuclei. The opacity of the universe consequently dropped, decoupling radiation from matter and providing an imprint of the last scattering surface that would be redshifted with the expansion of the universe. This redshifted remnant of the big bang was discovered in 1965 as isotropic radio noise detected at Bell Labs by Penzias and Wilson, only 2K cooler than predicted by Alpher and Gamow. Fred Hoyle's nucleosynthesis models, Martin Ryle's studies on the evolving radio galaxy population and cosmic-age estimates through Walter Baade's improvement on H_0 formed a body of evidence leading to general acceptance of the model.

1.3 The realm of Quantitative Cosmology

Our place and importance in the universe has gradually eroded from lead role on centre-stage, to supporting actor, and finally to but one of many participants in a gigantic theatre. The currently accepted description of the universe is comprised of two components: the General-Relativistic treatment of curved space (for a review, see Lahav and Suto, 2004) and models describing the physical content of the universe. Whilst these components are intertwined, in the very broadest sense the work presented in this thesis concerns the latter.

The modern model of the universe originates from the Cosmological Principle, requiring a cosmos both homogeneous and isotropic. From these assumptions and a simple Newtonian treatment[2], one can describe the expansion of the universe with the Friedmann equation:

$$H^2(t) = \frac{8\pi G}{3}\rho \; - \; \frac{kc^2}{a^2} \; + \; \frac{\Lambda}{3}, \tag{1.2}$$

where $H(t)$ is the Hubble parameter at time t, ρ and k describe the density and geometry

[2] see Liddle (2003) for a derivation

of the universe, a is the scale factor and Λ is the cosmological constant (see §1.3.3). For a flat (k=0) universe (and assuming $c = 1$) one can define a dimensionless, time-evolving matter-density parameter, measured relative to the critical density $\rho_{\rm c}$[3]:

$$\Omega_{\rm m}(t) = \frac{\rho_{\rm m}(t)}{\rho_c(t)} = \frac{8\pi G\rho_{\rm m}(t)}{3H^2(t)}, \tag{1.3}$$

and an analogous density parameter for the cosmological constant:

$$\Omega_\Lambda(t) = \frac{\Lambda}{3H^2(t)}. \tag{1.4}$$

A flat universe therefore requires the condition:

$$\Omega_\Lambda + \Omega_{\rm m} = 1. \tag{1.5}$$

One may further subdivide the matter density of the universe into Dark and Baryonic components ($\Omega_{\rm DM}$ and $\Omega_{\rm b}$ respectively), reflecting the observations detailed in §1.3.1 and §1.3.2.

Measurements of H_0 from the *Hubble Space Telescope* (Riess et al., 2009), Baryon Acoustic Oscillation measurements from the SDSS (see Percival et al., 2010, and §1.3.2 later) and acoustic peaks in the power spectrum of the Cosmic Microwave Background (CMB) from the *Wilkinson Microwave Anisotropy Probe* (*WMAP Jarosik et al., 2011)* have converged astronomical thinking to what is now considered a concordant model: the "Lambda Cold Dark Matter" (ΛCDM) universe. Partially depicted in Figure 1.1, this comprises a geometrically flat universe (Ω=1) with the Hubble Constant in eqn 1.1 estimated as 71km s^{-1} Mpc^{-1} (parametrised as h=0.71). The universe is dominated by dark energy (with Ω_Λ=0.72), followed by dark matter ($\Omega_{\rm DM}$=0.23) and baryonic matter contributing the remainder ($\Omega_{\rm b}$=0.05). The most striking result from these findings is that over 95% of the universe does not appear on the periodic table. Understanding the nature of the dark universe remains one of science's major challenges.

1.3.1 Baryons (ordinary matter)

The initial atomic mass distribution and amount of baryonic matter was set within the first twenty minutes of the big bang. Particles that interact with electromagnetic radiation are predominantly "matter" as opposed to "antimatter". This implies asymmetry

[3]This is the density required to halt the collapse of the universe. Current observational data suggest a geometrically flat universe very close to critical.

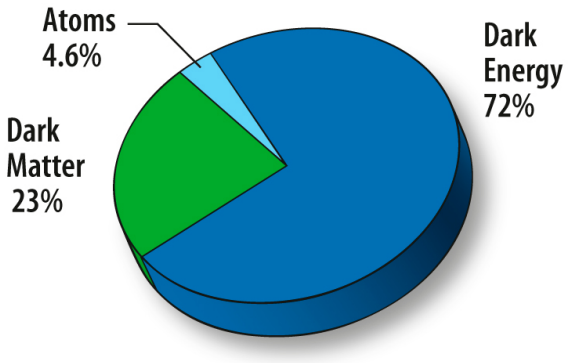

Figure 1.1: The composition of the universe, based on 5-year *WMAP* data (Komatsu et al., 2009). Less than 5% of the universe is on the periodic table. (source: `http://nasa.gov`)

in matter reactions and anti-matter reactions that preferentially select the former. In 1967 Andrei Sakharov proposed three conditions required to explain this break in symmetry, two of which are (or have been) verifiable[4]. Although the primordial plasma cooling rate primarily set the amount of Hydrogen, Helium, Lithium and (trace) Beryllium formed, the synthesis of heavier elements proceeded via the life-cycles of first-generation stars.

In observational cosmology, detected light may be used to trace the distribution and kinematics of baryons. Matter, in a diverse range of forms from interstellar dust grains to hot intra-cluster plasma, interacts electromagnetically through a wide range of mechanisms - one therefore requires detection capability over the full electromagnetic spectrum. Radiation emission and absorption processes will often provide clues as to the physical (density, temperature), magnetic and chemical state of the material observed. The fundamental challenge in cosmology is book-keeping: determining how much there is of a particular component, and how it has changed over cosmic time. Because one has a good handle on the exchange rate between light and mass for a variety of astronomical objects, observations of the local universe permitted estimates of how much mass there is in the whole universe[5]. The comparison between luminosity-inferred and dynamical masses first highlighted the cosmos was perhaps dealing on the black market.

[4]Under the standard model, violation of Baryon number may occur under circumstances. Both C & CP violation have been observed.

[5]With the advent of modern (in particular microwave) instrumentation, there are now alternative approaches to weighing the universe.

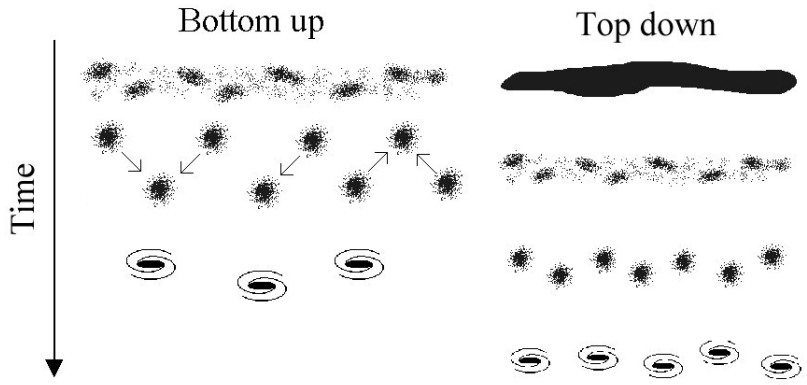

Figure 1.2: Two mechanisms for the formation of structure as the universe evolves. In the "bottom-up" (*left* panel) model, small structures coalesce to hierarchically construct larger structures. Conversely, with "top-down" (*right* panel) formation, the very largest structures formed first, subsequently fragmenting into smaller systems. Current observations and simulations suggest the "bottom-up" scenario is most likely.

1.3.2 Dark Matter

In 1937 Fritz Zwicky undertook a census of the Coma cluster (Zwicky, 1937), a virialised conglomeration of galaxies approximately $60h^{-1}\mathrm{Mpc}$ away. He compared the average dynamical mass of member galaxies (determined by the Virial theorem) to the luminosities of the galaxies. Unable to reconcile two orders of magnitude difference between these two estimates, Zwicky commented "a further analysis of the problem is in order".

Over the coming 50 years, this "problem" manifested itself in other astronomical investigations. Looking to individual galaxies, Vera Rubin's 21cm measurements of the radial rotational velocity in spiral galaxies produced non-Keplarian profiles. 21cm lines are capable of measuring neutral Hydrogen emission out to radii too faint to be probed by stellar emission. The observed flatness of the profile out to large radii suggested additional, unseen matter produces a uniform mass density to radii where the observed density is low. Although the influence of this unseen dark matter is clear in many spiral galaxies, the evidence in elliptical galaxies is somewhat more contentious (see Romanowsky et al. (2003) and Dekel et al. (2005) for contrasting outlooks).

Dark matter also accounts for the extent to which the structures we observe have collapsed. According to the currently understood picture of structure formation, the universe assembled "bottom-up". As illustrated in Figure 1.2, small structures initially collapse and merge to form compact structures sooner than the alternative "top-down" model. Throughout this process, the gas pressure acts against gravity. The short timescale over which initial structures formed suggests a significant mass component gravitationally assisted this collapse non-interactively: without dark matter, galaxy formation would begin at later times. The statistical clustering properties measured from dark-matter simulations match well the observed distribution of large scale structure over the history of the universe (e.g. Conroy et al., 2006).

Although smooth on large scales, original structures were seeded by small inhomogeneities evident in the CMB radiation field. Thought to arise from quantum-scale fluctuations frozen in after the rapid expansion of the universe at 10^{-36} seconds (inflation), the small-scale density anisotropies in the photon-baryon fluid were encoded in the radiation field at the point of decoupling.

Prior to the era of recombination acoustic oscillations permeated the fluid, driven by competition between baryonic gravitational compression and radiation pressure from the photons. These oscillations left imprints in the CMB, resulting in a series of acoustic peaks in the angular power spectrum of the radiation field. Related to these primordial sound waves is the origin of Baryon Acoustic Oscillations (BAO). At the point of decoupling, the universe becomes neutral, and photons free-stream out of the fluid. Without radiation pressure, baryons in the expanding spherical sound waves (caused by acoustic oscillations) collapse under their own gravity at the sound horizon and leave imprints on forming structures throughout the history of the universe. The scale-size of this horizon, well defined by theory, depends on the amount of baryonic ($\Omega_{\rm b}h^2$) and dark matter ($\Omega_{\rm DM}h^2$) in the universe. This means searches for a low-amplitude peak in the large-scale (r$\simeq$110h^{-1}Mpc) clustering of galaxies are at the forefront of cosmological research, with detections in the 2dFGRS (Cole et al., 2005) and SDSS (Eisenstein et al., 2005).

Whilst the last scattering surface of the CMB was baryonic matter, it was the dark matter that set the agenda for the distribution of mass that we see imprinted in Figure 1.3. This plot shows data from *WMAP* (Jarosik et al., 2011), the second (of three) satellites to image the entire sky at microwave frequencies. Secondary anisotropies arising from

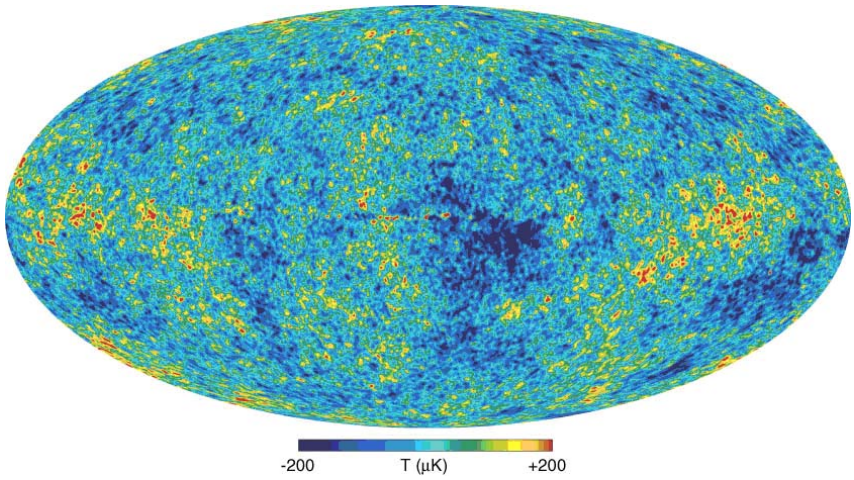

Figure 1.3: (from Hinshaw et al., 2009) The microKelvin-scale temperature deviations in the Cosmic Microwave Background radiation, as measured by the *WMAP* satellite. Blue colours indicate regions colder than the mean temperature, red are regions hotter than the mean temperature. The deviations from the mean are as small as 1 in 10^6.

our motion relative to the CMB rest frame have been corrected (in addition to contributions from the galactic plane), and micro-Kelvin deviations from the mean temperature have been colour-coded. The blue points correspond to regions of low temperature (likely seeding today's galaxies and clusters), whilst higher temperature red points will correspond to lower-density regions in the universe. The position and height of acoustic peaks in the power spectrum measured from the clustering of primary anisotropies[6] depend sensitively on the cosmology of the universe. By comparing this power spectrum to models with different cosmologies, data can be fitted to a ΛCDM model with high confidence. Figure 1.4 shows the measured power spectrum based on a range of CMB experiments covering different spatial scales.

Dark matter can also alter the local curvature of space. Significant accumulations of mass may act as gravitational lenses, capable of focussing and magnifying the light of distant objects. The strength of this effect depends on the configuration of source and lens, and the mass of the latter. Measurements of lensing due to galaxy clusters have

[6]Anisotropies in the CMB may arise from either signals at the time of last scattering (primary) or subsequent modifications to this signal (secondary, or late-time anisotropies) as the radiation interacts with the evolving universe.

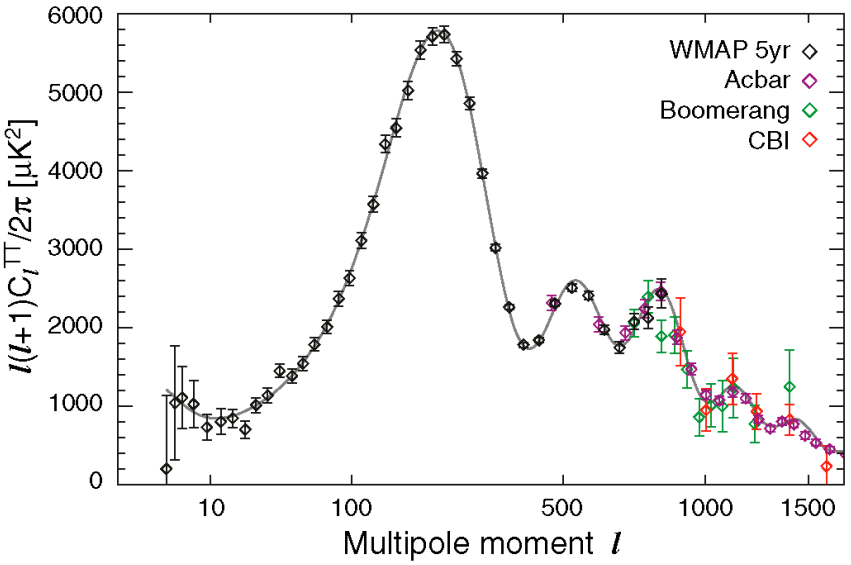

Figure 1.4: (from Nolta et al., 2009) The CMB power spectrum from the ΛCDM model and measurements from *WMAP* (satellite), Boomerang (balloon-borne Jones et al., 2006), ground-based CBI and ACBAR (Readhead et al., 2004; Reichardt et al., 2009) data. The geometry and composition of the universe influences the position and amplitude of the acoustic oscillations.

revealed considerably more mass is required than observed to explain the degree and frequency of lensing. Dark matter has been directly probed by combined lensing, X-ray and optical studies of the "Bullet Cluster" (Clowe et al., 2006). Figure 1.5 shows this merger of two clusters. The dominant baryonic mass components (hot intra-cluster gas, pink) in each cluster have become detached from the collisionless dark matter haloes (and quasi-collisionless galaxies). Weak lensing studies of the region note the majority of mass resides with the collisionless component of the merger (*blue*). Galaxies on their own have insufficient mass to account for the observed level of lensing, and it is instead the dark matter haloes that produce the signal.

1.3.3 Dark Energy

Over the last few years, astronomical experiments have supplied strong evidence for an increasing rate of expansion in the universe. Within the context of the standard cosmological model, some 72% of the matter-energy budget consists of a dark energy opposing

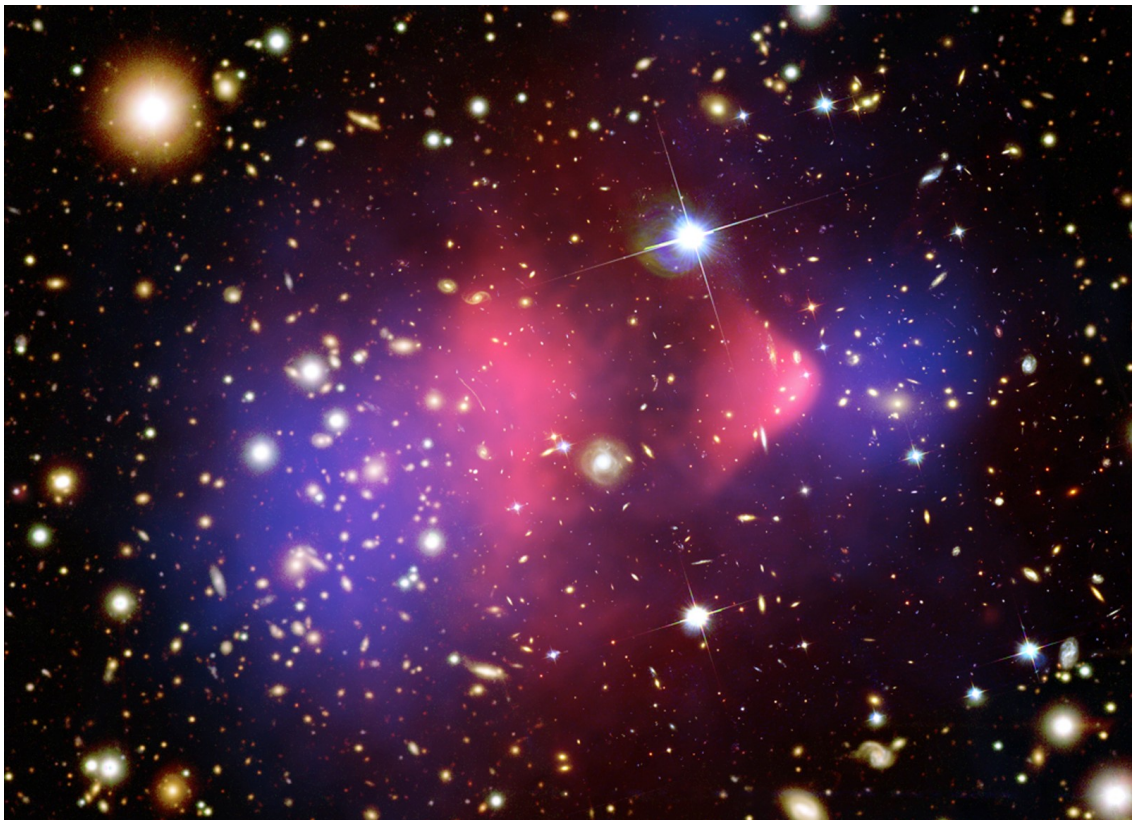

Figure 1.5: The bullet cluster (1E 0657-56). In fact two merging systems, this composite image shows the location of the ICM (pink, X-ray Markevitch, 2006) and weak-lensing inferred dark matter (blue Clowe et al., 2006). (source: `http://chandra.harvard.edu`)

the effect of gravity and accelerating the cosmos.

Evidence suggesting an accelerating universe came originally from calibration studies of high redshift Type 1a supernovae (SN). Galaxy redshifts and supernovae are two distance indicators - commonly known as "standard candles". Where two standard candles overlap, their results may be calibrated to ensure precision and consistency. Although there is a good agreement between the distance implied by redshifts and those by supernovae locally, as one looks to greater redshifts, the relation becomes non-linear. Distant ($z \gtrsim 1$) SN appear approximately $1/4$ magnitude fainter than expected from a decelerating universe. Fitting data from two independent high-z SN searches (Perlmutter et al., 1998; Riess et al., 1998) to a Λ-based cosmological model strongly suggests a non-zero cosmological constant. In the broadest of terms, dark energy is a cosmological *terra incognita* encapsulating current ignorance of the underlying mechanism. Two

leading dark energy candidates are Einstein's cosmological constant, and exotic forms of scalar field termed quintessence. A vacuum energy pervading the universe, functionally equivalent to a cosmological constant, may accelerate the universe without the need for additional physics. The considerable tension between estimated vacuum energy densities and those predicted by Quantum Field Theory (see Carroll, 2001; Frieman et al., 2008) may be eased with the introduction of supersymmetric particles currently being searched for by the Large Hadron Collider at CERN. With experimental verification of supersymmetry, one may potentially (and elegantly) account for both dark matter (for which the lightest supersymmetric particle is a candidate) and dark energy simultaneously (Weinberg, 1989). Other alternatives include a breakdown of General Relativity and failures of the standard cosmological model. The case for dark energy, as with dark matter, is strongly inductive: empirical observations were not motivated by the formulation of a hypothesis grounded in known physics. Despite this, dark energy is considered one of the most important cosmological discoveries and has profound implications for many branches of Physics.

Originally Einstein's cosmological constant, Λ, was invoked to counteract the expansion of the universe when it was thought to be static. We now use this term to describe an energy density pervading the universe that does not vary with time or position. Such a scenario arises from a dark energy equation of state (relating the pressure P to the density ρ) $w = \mathrm{P}/\rho = -1$. Quintessence-type solutions permit time and potentially position-dependent equations of state: $w = \mathrm{P}/\rho = w(\mathrm{z}, \mathbf{r})$. Galaxy clusters can discriminate between these models by measuring the evolution of their population with cosmic time. Such evolution is sensitive to dark energy through the growth of the volume with time, and the competing effects of dark energy and gravity.

1.4 Observing the large-scale universe

The study of large scale structure is essential to unlocking the nature of the universe. As discussed above it can infer the ingredients making up the universe, probe the formation of structure, and explain the role nature and nurture play in the evolution of galaxies. Observationally, galaxies are the basic building blocks of the megaparsec-scale universe, but in reality it is the dark matter haloes hosting them. Studying the distribution and

evolution of matter over very large volumes permits constraints on parameters describing the universe. One key ingredient to the detection, for example, of BAOs is a large survey volume. Probing non-gaussianity and higher-order clustering similarly requires large volumes to ensure statistical reliability. Furthermore, topological studies of the universe are only possible when the sample is considerably larger than the scale size of the features studied. Among many other drivers, these steer cosmological research towards gigaparsec-scale surveys, scanning both area and redshift to ensure simultaneous coverage of evolutionary domain (depth) and sample size (breadth).

1.4.1 Filaments

Although clusters appear to be the largest virialised systems in the universe, the assembly of structure does not necessarily stop there. Despite the Cosmological principle calling for homogeneity at large scales, the scale itself is somewhat ambiguous. Of particular interest therefore is how mass is structured and connected.

Through studies of large 3D maps of the universe and dark matter simulations, one is able to characterise the large-scale distribution of light. There has been considerable progress in explorations of both real and simulated universes. The former, fuelled by improvements to electronic detectors such as CCDs, has allowed the detection of millions of galaxies in the optical and near-IR through surveys such as the Sloan Digital Sky Survey (SDSS York et al., 2000) (SDSS; York et al., 2000), Two-degree Field Galaxy Redshift Survey (2dFGRS; Colless et al., 2001) and 2MASS (Skrutskie et al., 2006). Accurate cosmographic maps detailing the distribution of galaxies in these surveys provide a useful basis against which models of structure formation at the largest scales can be tested.

Modern simulations of the universe have exploited the significant advances in computational hardware and a better knowledge of the initial conditions (Jenkins et al., 1998) that seeded structure in the universe. Vast N-body simulations from teams such as the Virgo consortium[7] describe how dark matter accumulates, and show filamentary channels feeding into dark matter haloes acting as the hosts for galaxies and clusters of galaxies. The focus of these large realisations shifts, depending on the investigation, between large-volume rendering (Millennium; Springel et al., 2005), selected re-simulations at

[7] http://www.virgo.dur.ac.uk

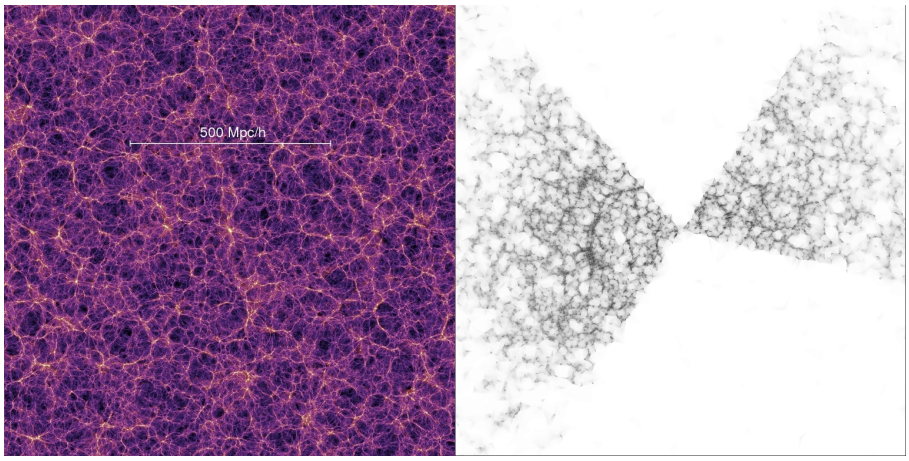

Figure 1.6: The simulated (*left*, Millennium Simulation Springel et al., 2005) and observed (*right*, The Sloan Digital Sky Survey York et al., 2000) cosmic web. The Millennium Simulation data shows the distribution of dark matter, whilst the survey data shows smoothed light from galaxies. Both display distinct filamentary structure, suggesting broadly an agreement between theory and data. (source: Virgo Consortium & R. van de Weijgaert)

high resolution of lower resolution data (GIMIC; Crain et al., 2009) and galaxy-scale simulations (Aquarius; Springel et al., 2008).

At megaparsec scales, matter is structured in what has been coined the cosmic web: a connected sponge-like system with clusters at high-density vertices, interconnected by filamentary threads of galaxies tracing the underlying dark matter network. Both these filaments and vast sheets of galaxies serve to delineate the large voids containing very few galaxies. This cosmic foam is highly anisotropic - for example, matter is funnelled onto cluster infall regions connected to filaments. Figure 1.6 shows the observed cosmic structure in the Millennium Simulation and the smoothed galaxy distribution from the SDSS. Despite visual impressions from this Figure, it remains a challenge to define filaments. They are commonly thought of as connected systems of galaxies between clusters, precise characterisation remains elusive and subjective.

The study of large-scale structure has been made possible by large bodies of data, typically in the form of cluster and galaxy catalogues. The three catalogues of most historical importance were the Abell (1958) catalogue, a series of compilations by Fritz

Zwicky (the first volume being Zwicky et al., 1961) and the Shane and Wirtanen (1967) catalogues. The latter consisted of one million galaxies counted from Lick Observatory plates. These were later analysed by Seldner et al. (1977) to determine the angular distribution of the galaxies. These catalogues formed the basis for many surveys, in which the angular (and in some cases 3D) galaxy distributions were investigated. Abell noted in the abstract of his work that there was evidence of "second-order clusters" - a hierarchical structure in the universe, suggesting a "clustering of clusters"; indeed de Vaucouleurs had previously recognised the Virgo supercluster (Rubin, 1951; de Vaucouleurs, 1953a), calling it a "supergalaxy".

Real progress came with early redshift catalogues, where the 3D distribution of galaxies not only enabled the discovery of superclusters and filaments of galaxies, but also provided an observational basis for competing structure formation scenarios. One survey particularly worthy of note was the Harvard CfA redshift survey (de Lapparent, Geller, and Huchra, 1986). This led to the production of one of the more distinctive galaxy distribution maps, and clearly highlighted the deviation from a smooth distribution of galaxies. The central "stick man" feature, existing as part of a larger structure later termed the *Great Wall* (at cz$\simeq$10,000kms^{-1}) clearly demonstrated that the universe is "foamy". As galaxy redshift surveys increased in scale, the true extent of how galaxies are distributed became apparent. In particular the 2dFGRS[8] (Colless et al., 2001) observational campaign, amassing approximately 250,000 galaxies, provided a reliable picture of the universe out to a redshift of 0.12. Studies to uncover the large-scale structure within this survey have identified both voids (Croton et al., 2004a) and filaments (Pimbblet et al., 2004). By blurring the distribution of galaxies over a smoothing scale of 3 and 15h^{-1}Mpc, Baugh et al. (2004) also highlighted two particularly large collections of galaxies in this survey (seen in Figure 1.7). It remains a challenge for cosmological models to produce the correct clustering signal over the entire dynamic range of scales galaxies are seen to associate. A comprehensive comparison between model and observed structures has yet to be performed, with efforts typically focused on the characterisation of filaments in dark matter simulations (e.g. Colberg, 2007).

Studies of individual filaments within ΛCDM simulations (e.g. Colberg et al., 1999) suggest their gas and galaxies may contribute up to 50% (Davé et al. 2001, Cen and Os-

[8]Two-degree Field Galaxy Redshift Survey, `http://www.mso.anu.edu.au/2dFGRS`

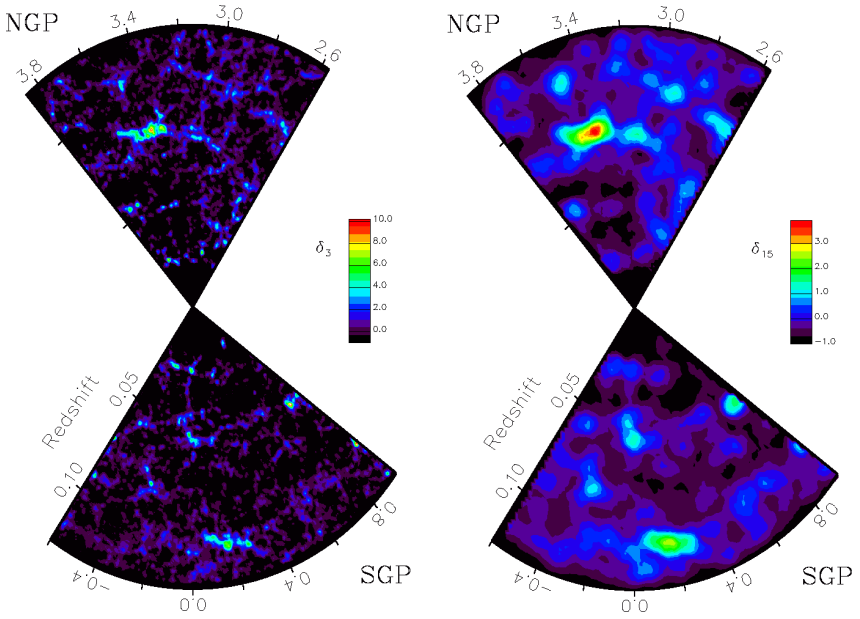

Figure 1.7: (from Baugh et al., 2004) Superclusters in the RA-z plane of the 2dFGRS. The 3D positions of galaxies are smoothed on scales of $3h^{-1}$Mpc (*left*) and $15h^{-1}$Mpc (*right*). These maps highlight significant galaxy concentrations within the North Galactic Pole (NGP) at z$\simeq$0.07 and the South Galactic Pole (SGP) at z$\simeq$0.11.

triker 2006) of the cosmic baryonic mass budget in the form of a Warm-Hot Intergalactic Medium (WHIM). This is supported by observations of the local Intergalactic Medium (Penton et al. 2004), suggesting the filamentary WHIM component plays host to a significant fraction of the directly observable universe. However, this is only sparsely sampled by the galaxy population, meaning significant challenges arise from the direct identification of filaments. Tentative evidence of the gas component through soft-band X-ray detections (Scharf et al., 2000; Werner et al., 2008) have been subsequently supported by weak lensing shear induced by the mass of the filaments (Dietrich et al., 2005). Future radio surveys such as ASKAP's WALLABY[9] also hold great potential for the identification of HI-selected filaments. However, from the broader 100-megaparsec perspective of filamentary structure, it is galaxy redshift surveys that facilitate large-scale cosmography.

[9] `http://www.atnf.csiro.au/research/WALLABY`

1.4.2 Groups & Clusters of galaxies

In the hierarchical picture of structure formation, the bottom-up merging of dark matter haloes will inevitably result in the assembly of virialised collections of galaxies. These may range from groups of a handful of galaxies, through to vast 1000-member clusters such as Coma. The Milky Way and our two nearest neighbours, the Small and Large Magellanic Clouds are members of a 20-galaxy association called the *Local Group*. This group is attracted to the Virgo cluster some $10h^{-1}\mathrm{Mpc}$ distant.

For a typical cluster, the majority (over 90%) of baryonic mass is in the form of a hot (10^7-10^8K) tenuous plasma pervading the system[10]. This *Intra-Cluster Medium* (ICM) can be directly detected with X-rays and indirectly through its imprint on the CMB (the Sunyaev-Zel'dovich effect), but is unobservable in the optical regime. Figure 1.8 shows Abell 2029 in X-ray (*left*) and optical light (*right*). The remaining baryonic mass comes in the form of the galaxies observed through optical studies, and it was through visible observations that early cluster detection was performed.

Up until the 1980s, very few galaxies had measured redshifts. It was therefore a significant feat to detect three-dimensional objects with only two dimensions. Abell's aforementioned 1958 work remains one of the most significant cluster maps in Astronomy. The catalogue recorded a total of 2712 clusters, using the visual inspection of Palomar Observatory Sky Survey plates.

Abell's catalogue was considered superior to that released by Zwicky, due to a more rigorous approach towards categorising and defining clusters. It remains at the forefront of astronomy as one of the most cited publications within the field. Abell's principal approach was to count the number of galaxies within a specific magnitude range, inside a certain radius from the centre. This *Abell radius* was determined in part through the established cluster luminosity function $\Phi(L)dL$.

Luminosity functions describe the mean number of galaxies per unit volume within the luminosity range [L, L+dL] and are well described by the Schechter (1976) function. Combining $\Phi(L)dL$ with the apparent magnitude of the tenth brightest galaxy (where $\Phi(L)dL$ is well characterised, and the chance of field contamination lower) yielded a cluster distance used to establish a $1.5h^{-1}\mathrm{Mpc}$ radius aperture. Within this aperture all

[10] rather than localised around member galaxies

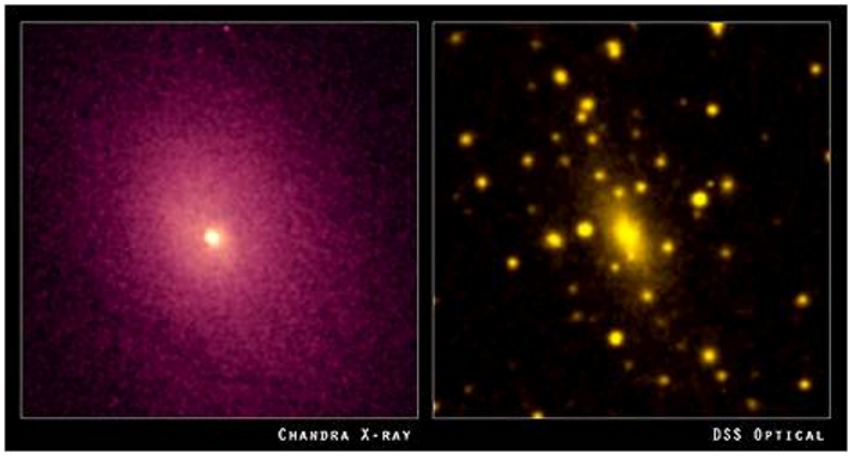

Figure 1.8: Two views of Abell 2029. The X-ray Chandra image ((*left*) Lewis et al., 2003) traces the ICM and is best suited to mass estimates of the clusters. Optical data (*right*, Palomar Observatory) resolves the galaxy members, and with spectroscopic data can determine the distance to the cluster. (source: `http://chandra.harvard.edu`)

cluster members between m_3+2 and m_3 were counted[11] and following a background correction to this count, clusters were divided into richness groups R. Although the groupings are somewhat arbitrary, they provide a good indication of the cluster mass; populations of R$\geq$1 groups are considered complete out to a redshift of z=0.08. However, concerns over the extent of projection errors and incompleteness in the poorest richness (R=0) group (see, for example Lucey, 1983) has limited the full scientific potential of this census. Spectroscopic campaigns targeting Abell clusters (e.g. Struble and Rood, 1987) allowed for such projection effects to be quantified and compensated for, permitting the determination of more precise clustering statistics (Sutherland, 1988). Armed with reliable catalogues of galaxy clusters, many studies have turned to individual clusters themselves, often combining measurements from deep imaging over a variety of wavelengths (such as Abell 2029 in Figure 1.8) and spectra of many cluster members.

A galactic census of any cluster is a study of gas-galaxy interactions and the impact of environment on galaxy properties. Perhaps most notable is the predominance of red elliptical galaxies with old stellar populations and trace star formation. That this was not always the case indicates the transformative process galaxies undergo when accreting

[11] m_3 denotes the third (apparent) brightest galaxy in the cluster

onto a cluster. Tidal harassment of captured disk galaxies strips them of the gas used to form stars, transforming them into passively evolving early-type galaxies.

The Butcher-Oemler effect (Butcher and Oemler, 1978) describes the increased fraction of blue, star-forming galaxies in high-redshift clusters relative to those observed locally. The observational verification of this effect remains contentious even today (see e.g. Andreon et al., 2006; Gerke et al., 2007, for contrasting views), primarily due to concerns of cluster contamination with blue field galaxies at similar redshifts, but additionally from the need to disentangle evolutionary effects from environmental (Metevier et al., 2000). Nevertheless, clusters do not exclusively play host to "red and dead" galaxies: some evidence (e.g Bildfell et al., 2008) suggests that cluster cores may exhibit some low-level star formation. Not only is the source of cold gas fuelling this formation disputed, but the role AGN activity plays in suppressing star formation in groups and clusters is only now becoming clear (e.g. McCarthy et al., 2011).

Although this thesis focuses on the optical selection of galaxy clusters, it is worth noting X-ray astronomy has in the past provided a rich source of cluster detections, complementary to optical surveys, through identification of the ICM. As discussed earlier, the vast fraction of cluster baryonic mass exists in the form of hot intracluster gas. The X-ray emissivity arising from thermal bremsstrahlung in this gas depends on the square of the electron density (Felten and Morrison, 1966). This property provides a useful observational characteristic with which to search for clusters of galaxies. Not only do extended X-ray emissions identify gas specifically from physically bound systems, their density dependence results in high-contrast cluster signals not subject to the projection effects found in optical imaging. Because X-ray cluster selection is sensitive to ICM physics, one can construct (subject to a proper understanding of the gas physics) flux-limited surveys with well-defined selection functions. This latter point, coupled with the strong correlation between X-ray luminosity and cluster mass, allows an exploration of the cosmologically-dependent evolution of these systems.

Satellite-based X-ray observatories such as *ROSAT*, *Chandra* and *XMM Newton* have been put to good use identifying and characterising clusters. The *ROSAT* all-sky survey (RASS; Voges et al., 1999) was the first all-sky X-ray census of the sky and spawned a number of X-ray cluster surveys including NORAS (Böhringer et al., 2000) and a 400 square degree high-galactic latitude survey (Burenin et al., 2007). The improved sensi-

tivity of *XMM Newton* and *Chandra* allowed for even deeper surveys and the selection of clusters beyond the redshifts detectable by optical surveys. In these instances, galaxy populations must be resolved by infrared facilities such as *Spitzer*, which are also capable of probing star formation in these systems (e.g. Hilton et al., 2010). The identification of a z=2.07 Virgo-mass cluster with extended X-ray emission (Gobat et al., 2011) has particularly significant implications for the assembly and virialisation timescale for distant clusters.

For each target selected by *XMM Newton* for long-exposure ($\sim$20ks) measurements, the surrounding area is simultaneously imaged. This has provided 5776 pointings that will be put to use in the forthcoming XMM Cluster Survey (XCS; Romer et al., 2001) for serendipitous cluster detections. The prescription for extracting clusters from approximately 400 square degrees of XMM imaging, detailed in Lloyd-Davies et al. (2010), has identified over 3700 cluster candidates, with over one third considered high significance (exceeding 300 counts). Sweeps of *Chandra* imaging have also been attempted (ChaMP; Barkhouse et al., 2006), but lower sensitivity and areal coverage (by factors 4.5 and 30 compared to the XCS) restricts the potential yield, despite better image resolution reducing the point source contamination.

Microwave detection of clusters is an emerging field and, somewhat akin to the X-ray approach, relies on gas residing in a large gravitational potential. Microwave telescopes such as the South Pole Telescope (SPT) and the Atacama Cosmology Telescope (ACT) search for the imprint of hot intracluster gas on the CMB. When passing through a cluster, approximately 1% of CMB photons encounter free ICM elections. Up-scattered by the inverse-Compton effect, these photons increase in frequency. This causes a detectable Sunyaev-Zel'dovich (SZ) decrement in the expected CMB signal: a secondary anisotropy correlated to the gas mass of the cluster. Blind searches of the sky with the SPT (Staniszewski et al., 2009) and ACT (Hincks et al., 2010) have recently begun, already uncovering distant SZ-selected clusters such as SPT-CL J0546-5345 (Figure 1.9), a z=1.067 cluster supported by deep *Spitzer* and *Chandra* imaging.

An entirely non-baryonic detection approach may be achieved with blind searches for the weak-lensing shear of background galaxies arising from the gravitational potential of the cluster. Studies such as those of Miyazaki et al. (2002) and Wittman et al. (2006) have used deep optical-band imaging to identify shear-selected clusters, subsequently

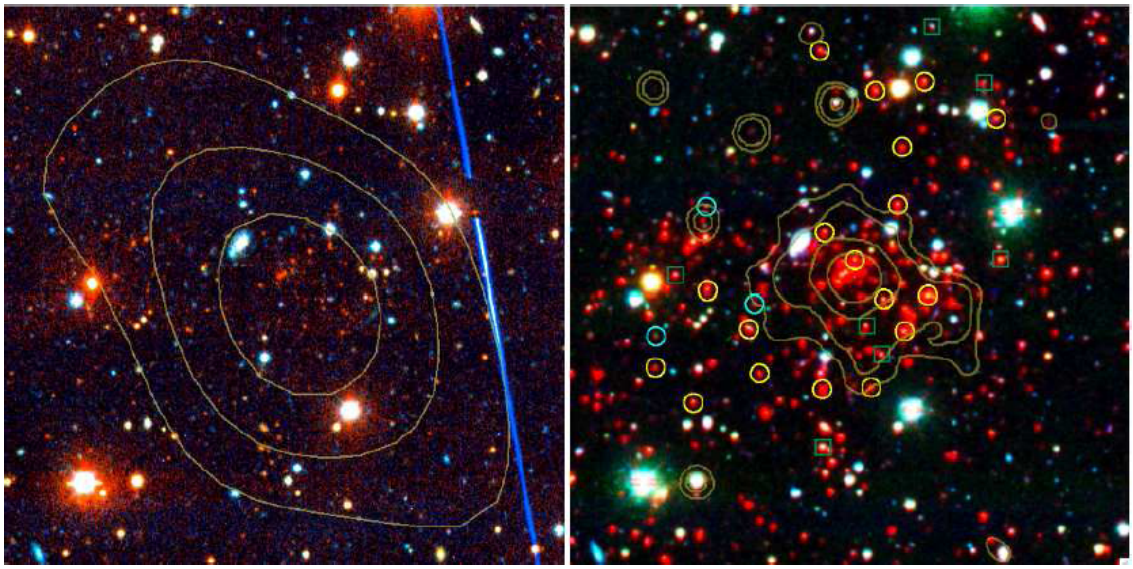

Figure 1.9: (from Brodwin et al., 2010) The z=1.067 SZ-decrement detected cluster SPT-CL J0546-5345. The *left* panel is optical *grz* imaging overlaid with SZ-decrement significance contours. The *right* panel is the same system in the *ri* and *Spitzer* 3.6μm bands with *Chandra* X-ray contours added.

verified by spectroscopic follow-ups. This method however requires high-quality photometric data and its application to wide-field surveys has yet to be tested.

Despite the advantages of SZ and X-ray cluster detection methods, they also have drawbacks. Both approaches are sensitive to the details (and so limited by our understanding) of intracluster gas physics. Early, unvirialised proto-clusters in the process of formation or those otherwise deficient in gas may remain undetected. Separating genuinely extended sources from point sources relies heavily on the beam size, with limits subsequently placed on the minimum detectable scale-size of clusters. Because SZ cluster identifications arise through absorption of the CMB, projection of discrete absorbing systems along the line of sight may also lead to false cluster signals.

Most important for the derivation of cluster masses are distance estimates. SZ decrements contain no redshift data, and although X-ray redshifts (identified through K-shell Fe-line searches) are possible (e.g. Yu et al., 2011), they suffer from low S/N and rely on the correct metallicity estimation and modelling of ICM physics. At the most superficial level, one cannot track the evolution and formation of galaxies in clusters without

resolving the population.

Technological and engineering advances have resulted in the renaissance of optical cluster selection. As optical redshifts are set to remain the most viable distance estimator, it is likely a holistic multi-wavelength approach to cluster identification will in the future yield the most productive science, with datasets providing complementary data on the same field.

1.4.3 Modern optical cluster and group detection

- Detection with spectroscopy

An important transition in modern cosmology was the move towards large, automated redshift surveys capable of sampling the spectra of multiple targets in a single field. Mentioned previously, the 2dFGRS was one particularly ambitious survey aiming to study 250,000 galaxies, followed closely by the Sloan Digital Sky Survey (SDSS) seeking redshifts for 1 million galaxies[12] (see Abazajian et al., 2009, for most recent progress).

With accurate three-dimensional redshift-space data available for all galaxies in a survey, the most effective means of connecting galaxies into groups is via the Friends-Of-Friends algorithm, first applied to astronomical data by Huchra and Geller (1982). Clusters are formed from galaxies connected within a common linking volume, defined either by a sphere or anisotropic (cigar / lozenge) linking volume (such as Tegmark et al., 2004). Application of a Friends-of-Friends algorithm to 2dFGRS data produced the 2dFGRS Percolation Inferred Galaxy Group (2PIGG) catalogue (Eke et al., 2004a). The catalogue was constructed by using a friends-of-friends algorithm that associated groups of galaxies by a linking length 13% of the mean galaxy separation. The catalogue identified over 29,000 clusters with at least 2 members, of which 7,000 had a richness of at least 4 members. The parameters chosen for the search were optimally selected from tests on mock galaxy surveys. In these simulated realisations of the survey, the galaxy membership of each dark matter halo is known. The recovery of galaxy groups can then be bench-marked against the true group population. Similar efforts have been attempted using spectroscopic data from the SDSS survey, such as Berlind et al. (2006) and Tago et al. (2008), but both surveys will suffer from a high-density incompleteness

[12] At time of writing, the number of unique spectra for galaxies in the SDSS stands at 976,839

arising from a minimum fibre separation (55″ for SDSS) that prevents measurement of two closely separated galaxies[13].

A number of large-scale spectroscopic surveys are now in progress. WiggleZ (Drinkwater et al., 2010) and SDSS3:BOSS[14] are both geared towards the detection of Baryon Acoustic Oscillations, the latter an extension to the SDSS campaign. Because they identify galaxies from [OII] emission lines, their samples tend to be blue, star-forming galaxies. Although evidence suggests such galaxies may be present in proto-clusters forming at high redshift (for example Capak et al., 2011), they are unlikely to trace the cluster population at the redshifts where their source densities permit reliable detections. The identification of both clusters and large-scale structure with these surveys therefore remains an interesting venture.

- Detection with angular and photometric data

The task of identifying clusters in optical surveys without redshift data is challenging. Although no photometric survey can hope to eliminate chance projections of galaxies, intelligent searches through these catalogues can improve the completeness while keeping spurious detections to a minimum. For many earlier surveys, only single-band data were available - as with Abell's catalogue, magnitudes were used as indicators of galaxy depth.

Using only angular data, one may smooth the galaxy positions using filters appropriate to the assumed projected distribution of galaxies within the cluster. Peaks in the smoothed field would therefore indicate potential cluster detections. Shectman (1985) use a very simple 3×3 kernel to identify overdensities in the Seldner et al. (1977) reprocessing of Shane and Wirtanen (1967) data. In order to refine this process, Gal et al. (2000) apply two kernels to their data: the first estimating the density field, the second changing size to account for variations in this field.

Whilst these algorithms search for cluster signals only in positional data, in principle one can supply additional constraints based on photometric data as well. Such "matched filters" can include priors based on the luminosity function and with two or more bands, colour as well. Postman et al. (1996) applies a matched filter sensitive to the evolving

[13] At $z\simeq0.1$, the median redshift of the survey, this translates to a separation of approximately $80h^{-1}\mathrm{Kpc}$

[14] `http://www.sdss3.org/surveys/boss.php`

cluster population, specifying radial distributions and luminosity functions designed to maximise a likelihood function. Matched filters can be even more elaborate, including cluster centroids constrained to the position of cD galaxies ("Brightest Cluster Galaxies" - BCGs common at the bottom of cluster-mass potential wells), photometric redshifts (see later) and similarities in cluster member colours.

Gladders and Yee (2000) were the first to use colour as the primary cluster selection criterion. Galaxies belonging to clusters are generally red early-types containing populations of uniformly old stellar populations. Bower et al. (1992) noted these stellar populations had been in place for over 2 Gyr in the Coma & Virgo clusters. The red colour results from a sharp break in the composite galaxy spectrum arising from strong (and age-dependent) CaII absorption lines[15] at 3933Å and 3968Å. CaII is common to cool F, G and K stars and becomes a progressively stronger absorber with stellar age. This "4000Å-break" in the galaxy spectrum can therefore be exploited when searching for clusters. The uniformity of cluster galaxy colours over a wide range in luminosities produces a distinct ridgeline in the colour-magnitude diagram known as the "red sequence" or colour-magnitude relation (CMR). The Gladders and Yee (2000) algorithm searched through colour-magnitude space in colour slices designed to isolate clusters at redshifts predicted by models of the red sequence. Both magnitude and radial density filters were built into the search procedure, but the authors noted the form of the density kernel does not significantly impact cluster detectability.

In an extension to this detection philosophy, Koester et al. (2007b) demonstrate the `maxBCG` algorithm on SDSS photometric data. This code searches through the catalogue and uses a likelihood function to rank galaxies according to how likely they are cluster BCGs. Members are assigned to highest ranking BCG candidates based on similarities in colour, luminosity and projected position.

The likelihood function describes the quality of match between putative detections and a model cluster with an NFW (Navarro, Frenk, and White, 1997) density profile, members exhibiting a red sequence, and the presence of the bight central galaxy. In measuring the peak response of the detection to this function, cluster redshifts and richnesses are calculated as part of the identification process. By imposing many model-

[15]These are traditionally known as the Fraunhofer H and K lines, with the suppression often called the HK-break

dependent constraints, concerns arise that clusters not conforming to the filter conditions are detected with lower significance. In the `maxBCG` case, the absence of a central cluster galaxy, or indeed the presence of a peripheral, bright galaxy may impact the effectiveness of the search (the latter instance arising from an assumed density profile radiating from a non-central location). By contrast, the `C4` algorithm (Miller et al., 2005) does not place as many constraints on the properties of systems it attempts to detect, instead searching for overdensities in a 7-dimensional space. Without directly specifying cluster properties or member-galaxy colours, `C4` is sensitive to a wider range of cluster types.

Building on the successes of the SDSS, wide-area multi-band optical photometric surveys are set to generate a wealth of astronomical data in the coming years. The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS[16]) prototype telescope, currently in operation on Haleakala, Hawaii, uses a 1.4 Gigapixel camera to image the entire visible sky in five bands several times per dark lunar cycle. The PS-1 prototype, currently generating between 3-5TB of data per night, will be extended in 2015-2016 to an array of four telescopes (PS-4) collectively producing over 10Tb per night. The Large Synoptic Survey Telescope (LSST[17]), due first light at approximately the same time, will be an 8m-class telescope with a 10deg^2 field of view imaged by 3.2 Gigapixel camera. Over a five year period, it is expected LSST will archive 5PB of catalogue data derived from 30PB of stored imaging. The work presented here explores early data from the PS-1 survey, and demonstrates how one might use photometric data to identify large-scale structure. Photometric redshifts are one promising means of exploiting this multi-band photometry to de-project the galaxy population.

- Detection with photometric redshifts

Spectroscopic surveys such as the 2dFGRS were designed as targeted follow-ups of sources derived from angular source catalogues. For one to probe fainter and hence deeper, longer-exposure source catalogues are required. One must then also consider, in the context of LSS recovery, the time-cost of recovering spectra for photometrically extended sources in any new catalogue. As one probes further down the luminosity function, the number of sources rises. Obtaining a high level of redshift completeness in a wide-area, deep survey is beyond the limits of current astronomical technique, prin-

[16] `http://pan-starrs.ifa.hawaii.edu`
[17] `http://www.lsst.org/lsst`

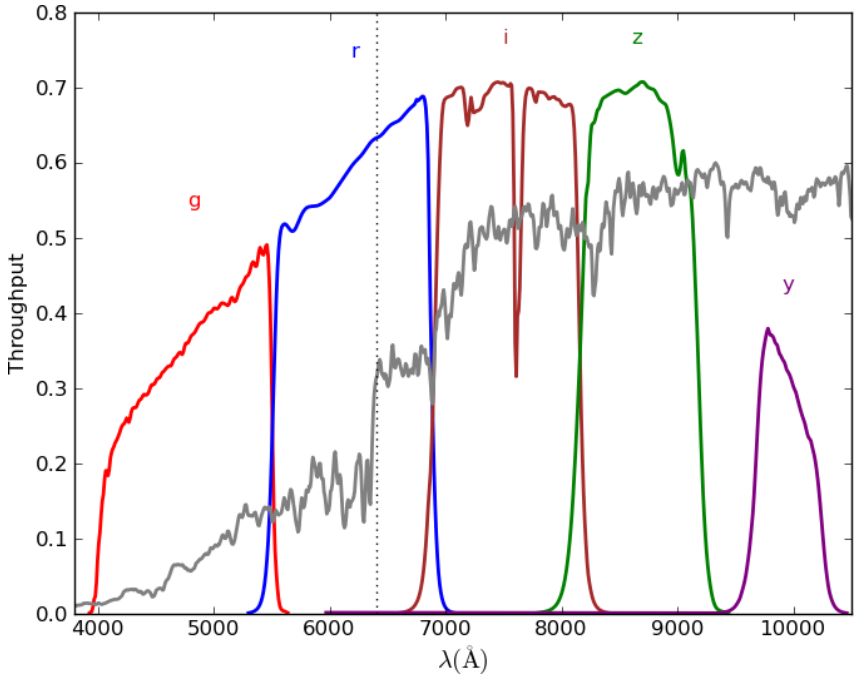

Figure 1.10: The simulated spectrum of a redshift 0.6 early-type galaxy. The position of the 4000Å feature is shown by the vertical dotted line. By observing this galaxy through a series of filters (the Pan-STARRS *grizy* optical system is included here), one may sample the spectrum of the source.

cipally constrained by the increased integration times (longer to attain sufficient S/N for fainter objects), source densities (not only limited by the available number of fibres, but also the non-zero separation between them) coupled with the proposed area of the survey (the 2dFGRS covers 3.5% of the full sky).

How else can we build a census of the local universe? As the sensitivity of detectors improved, and fields of view increased, astronomers looked to broad-band photometric imaging of the sky as a means to sample the universe. While this is not a new art, improved CCD sensitivities and spectral ranges have made this an attractive means of gathering digital data of the sky. In its simplest form, this involves merely taking a wide-field image through a filter. By taking repeat images of the same area, exposures can be stacked, and one can probe deeper still[18]. The selection of an object through this filter corresponds to a measurement of the observed spectral flux at the wavelength

[18]Searching for differences in the successive images is also fertile ground for the discovery of transients, supernovae etc.

corresponding to the filter, with the level of precision determined by the bandwidth of the filter (in the ideal case, neglecting instrumental effects such as quantum efficiency and spectral sensitivity). By measuring the flux of this same object in another filter, a different part of the observed spectrum is measured.

With many filters, one is able to well sample the spectrum of the source, as illustrated in Figure 1.10. Indeed spectroscopy could be thought of as the limiting case in this endeavour, where infinitesimal-width filters sample the spectrum over the observable range. The principle of photometric redshift estimation commonly relies on matching the coarse sampling of a galaxy's Spectral Energy Distribution (SED) to fluxes predicted from a series of template SEDs. Templates can be generated either from a model (such as Coleman et al., 1980) or alternatively a series of spectra from the survey itself (Csabai et al., 2003). Figure 1.11 shows template spectra taken from the Coleman et al. (1980) study, and illustrates the principle of sampling the SED with broad-band filters. In this instance, we have overlaid positions of the SDSS *ugriz* filter systems onto $\mathrm{z} = 0$ model spectra. The legend indicates the Elliptical, Irregular and two Spiral models (respectively E, Im, Sbc, Scd). In this example, the u and g bands offer the best resolving power between the different templates.

To estimate the source redshift, templates are manually redshifted over a range to provide a series of sample spectra that one may fit the galaxy to. The best match, typically determined by a minimised χ^2, then establishes the photometric redshift of the galaxy:

$$\chi^2(\mathrm{z}) = \sum_{\mathrm{i}=1}^{\mathrm{N_{filters}}} \left[\frac{\mathcal{F}_{obs,\mathrm{i}} - \phi \mathcal{F}_{\mathrm{SED}}(\mathrm{z})}{\sigma_{\mathrm{i}}} \right]^2 \tag{1.6}$$

where $\mathcal{F}_{obs,\mathrm{i}}$ and $\mathcal{F}_{\mathrm{SED}}(\mathrm{z})$ are the observed and model-template estimated fluxes for the i^{th} filter. ϕ is a scaling constant and σ_{i} is the measured flux uncertainty in the i^{th} filter. There are limitations to this technique, however. The templates used must represent the spectrum of galaxy types observed; ignorance as to the nature of the target will lead to poor fits. The process of redshifting the model spectra, in its simplest implementation, does not take into account the evolution of the source with redshift. Furthermore, χ^2-fits to the data run the risk of converging to local minima and hence false solutions. In light of both the above and growing use of photometric redshift estimates in large-area surveys, improved techniques have been developed. These include coupling neural-nets

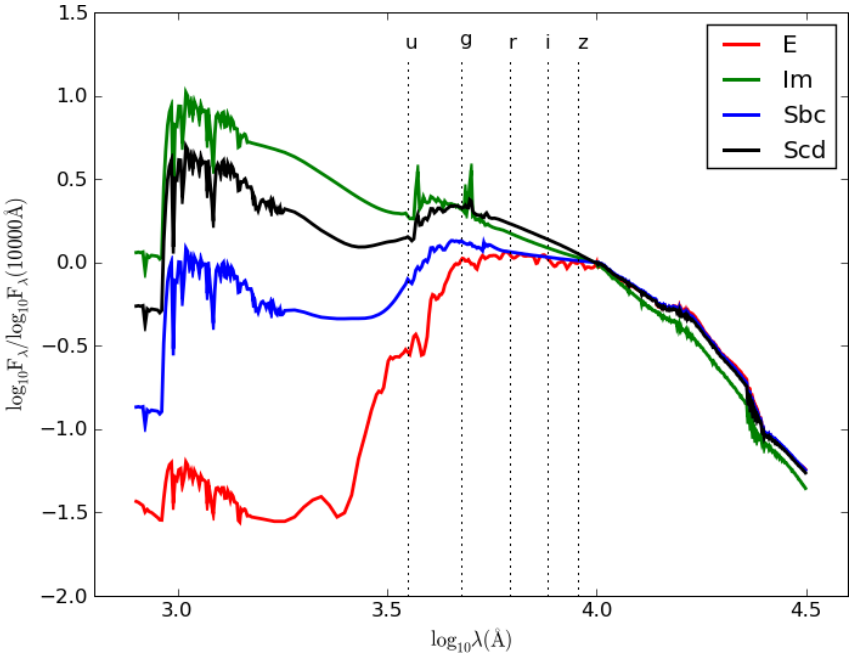

Figure 1.11: Model $z = 0$ templates from the Coleman et al. (1980) population synthesis libraries. These spectra have been normalised at 10,000Å and demonstrate how certain optical bands (such as the SDSS filter positions shown as vertical dotted lines) can discriminate between different galaxy templates.

with large training sets (Collister and Lahav, 2004) to machine-learning and template fitting hybrids (Abazajian et al., 2009, §4.6).

1.5 This work

The study of large-scale structure is an important tool in our efforts to test the ΛCDM concordance model, constrain models of galaxy formation and development, and understand the evolution of structure. In this thesis we present work that builds on previous efforts to identify and characterise large scale structure from deep and wide astronomical surveys. Because new observational campaigns will significantly increase the volume of data available for analysis, we demonstrate approaches that may be used to exploit these new data.

We start in chapter 2 with the recovery of connected galaxy structures using spectroscopic data. A cosmographic census of the connected nature of galaxies provides an insight into the higher-order clustering of mass, and probes the extreme scales describing the cosmic web. Using our prescription for identifying filamentary systems in galaxy surveys, we compare our detections to a series of realistic mock surveys. This allows us to establish how well the ΛCDM model recreates the large-scale distribution of galaxies, in particular superstructures of the size observed in the real data. We later use estimates of photometric redshift errors to simulate reconstruction of LSS without spectroscopy.

In chapter 3, we concentrate on the building blocks of these filamentary connected structures: galaxy clusters. Galaxy clusters are important for studying the dark sector of the universe. The evolution of both cluster gas and galaxies probes dark matter, whilst counts of the clusters themselves can constrain the nature of dark energy. In order to extract cluster data from forthcoming deep, multi-band wide-area surveys without the benefit of spectroscopy, the issue of de-projecting the galaxy population is addressed. We introduce a new cluster identification approach that exploits the colour properties and projected overdensity of their members, but otherwise makes only minimal assumptions. Our prescription is free from models that describe the clusters themselves, or the assumed spectral energy distribution of their members. Therefore, the cluster catalogues generated from this algorithm are complementary to matched filter and photometric-redshift approaches applied to the same data. To compare the recovery of clusters here with other algorithms, we perform a series of tests with mock data to gauge the efficacy of our approach.

We next apply this algorithm to a large optical survey. Because of the high volume of data under consideration, and the large parameter space explored, the computational implications of cluster searches in such a survey should not be ignored. In chapter 4, we discuss aspects of how the algorithm has been developed and improved to handle large datasets in preparation for very large surveys such as the Pan-STARRS 3π survey. Using a galaxy catalogue produced from the SDSS Stripe 82 survey, we present a catalogue of clusters detected over $270\mathrm{deg}^2$ of the sky. By post-processing cluster detections with available spectroscopic and photometric redshift data, we are able to calculate the cluster redshift distribution. Exploiting both accurate cluster redshift estimates and precise photometry from this large dataset, one may derive quantities to compare to model

predictions. One such observable, the redshift evolution of the colour-magnitude slope, provides a test of the mass-metallicity relation in simulations and permits refinements to subsequent cluster detections. We compare our estimate of sequence slopes to mock data generated from semi-analytic models, and later use our findings for a much larger survey.

Chapter 5 provides a snapshot of work in progress, and comprises a series of case-studies on galaxy surveys. Through these studies, we highlight additional development work that extends the capabilities of the detection algorithm. This culminates in the detection of clusters over an area over 30 times that of the study in chapter 3, as well as efforts to identify clusters beyond the limits of optical astronomy. Taking the largest survey studied, we provide cluster yield estimates for the next-generation of wide-area surveys and demonstrate the recovery of large-scale structures from these data in a similar manner to that of chapter 2.

Finally, we investigate one aspect of the cluster detection process: the characterisation of projected overdensities with Voronoi diagrams. A popular model used to describe the statistical significance of clustered distributions is known to be inaccurate in high-density environments most important to galaxy cluster detection. In chapter 6, we repeat the original study, verifying this inadequacy in the model. We go on to suggest an improved model that better describes the characterisation of random data with Voronoi diagrams. This new model may have potential application in cluster-finding and other astronomical studies.

Unless otherwise stated, throughout we assume a ΛCDM cosmology with $\Omega_{\rm m}$=0.3, Ω_Λ=0.7, H_0=70 km s^{-1} Mpc^{-1} and $h = H_0/100$km s^{-1} Mpc^{-1}. Unless specified otherwise, we use the AB magnitude system; for SDSS data we use the Sloan photometric system (Gunn et al., 1998) with model magnitudes.

Chapter 2

Detecting filamentary and connected structures in galaxy redshift surveys

2.1 Overview

In this chapter we describe and apply a simple prescription for defining connected structures in galaxy redshift surveys. The method is based upon two passes with a friends-of-friends groupfinder. The first pass uses a cylindrical linking volume to find galaxy groups and clusters, in order to account for the line-of-sight smearing introduced by the large random velocities of galaxies within these deep potential wells. The second pass, performed with a spherical linking volume, identifies the connected components. This algorithm has been applied to the 2dFGRS, within which it picks out a total of 7,603 systems containing at least two galaxies and having a mean redshift less than 0.12. Connected systems with many members appear filamentary in nature, and the algorithm recovers two particularly large filaments within the 2dFGRS. For comparison, the algorithm has also been applied to ΛCDM mock galaxy surveys. While the model population of such systems is broadly similar to that in the 2dFGRS, it does not generally contain such extremely large structures. When applying the algorithm to data perturbed to the uncertainty level of photometric redshifts, we find current estimators must reduce errors by at least an order of magnitude to recover the underlying structures.

2.2 Introduction

At very large scales, baryonic material is concentrated into an interconnected sponge-like network known as the Cosmic Web (Bond et al., 1996). Successive redshift surveys have traced out imposing overdensities in the galaxy distribution. Notable examples are the CfA 'Great Wall' (Geller and Huchra, 1989) and the "*Sloan Great Wall*" (Gott et al., 2005), which was also noted, if not named, in the Two-degree Field Galaxy Redshift Survey (2dFGRS, Colless et al., 2001) by Baugh et al. (2004) and Erdoğdu et al. (2004).

Most studies of large-scale structure in the Universe concentrate on measuring the galaxy power spectrum (e.g. Cole et al., 2005) or its Fourier transform, the two-point correlation function (e.g. Zehavi et al., 2002). These quantities will provide a complete statistical description of the galaxy distribution provided that their number density fluctuations are Gaussian. The standard model of structure formation, ΛCDM, assumes the initial inhomogeneities in the density field, generated during inflation, are similarly Gaussian. However, the subsequent growth of structure due to gravitational instability induces significant non-Gaussianities in the density field, and higher order moments of the density distribution become important. These phase correlations within the density field can be characterised either through higher order galaxy correlation functions (Baugh et al., 2004; Croton et al., 2004b; Gaztañaga et al., 2005; Nichol et al., 2006) or through the properties and distribution of filaments. Quantitative studies of filamentary structures in redshift surveys have been performed by Bhavsar and Barrow (1983); Barrow et al. (1985); Einasto et al. (2003); Pimbblet and Drinkwater (2004) and Stoica et al. (2010).

A variety of algorithms have been designed to describe the morphology of these large-scale structures using different techniques such as percolation (Bhavsar and Barrow, 1983), visual identification of regions between clusters (Pimbblet et al., 2004; Colberg et al., 2005), minimal spanning trees (Barrow et al., 1985; Colberg, 2007), the density field's Hessian (Aragón-Calvo et al., 2007; Bond et al., 2010; Zhang et al., 2009), gradient (Morse theory, Novikov et al., 2006) or linkage between the two (Sousbie et al., 2008a,b), the Hessian of the potential field (Hahn et al., 2007; Forero-Romero et al., 2009), Delaunay tessellations (Schaap and van de Weygaert, 2000; Aragón-Calvo et al., 2007), the Candy and Bisous models (Stoica et al., 2005, 2010) and watershed transforms (Sousbie et al.,

2009; Aragón-Calvo et al., 2010). Many of these algorithms also partition the whole of space into clusters, walls, filaments and voids. They are often applied to dark matter simulations to help describe the mass distribution, but with a few notable exceptions (Bhavsar and Barrow, 1983; Barrow et al., 1985; Pimbblet and Drinkwater, 2004; Bond et al., 2010; Stoica et al., 2010), they rarely include a comparison with observational data. A primary motivation for this chapter is to carry out a detailed quantitative comparison of filament properties using the 2dFGRS and mock galaxy catalogues, created using a ΛCDM simulation (Angulo et al., 2008) and a semi-analytical galaxy formation model (Baugh et al., 2005).

An important aspect of the comparison between model and observed large-scale structure concerns the existence in both the 2dFGRS and the Sloan Digital Sky Survey (SDSS; York et al., 2000) of some extremely large structures. Within the 2dFGRS (Colless et al., 2001) are two structures of particularly significant luminosity and extent (e.g. Baugh et al., 2004; Erdoğdu et al., 2004; Eke et al., 2004a). The structure in the northern slice of this survey has a redshift of $z \simeq 0.08$ and forms a part of the aforementioned *Sloan Great Wall*. The structure in the south appears more extended and is found at a redshift of $z \simeq 0.11$. These objects are known to have a large impact on the higher order correlations of the galaxy distribution (Baugh et al., 2004; Croton et al., 2004b; Gaztañaga et al., 2005; Nichol et al., 2006), and on its topology (Park et al., 2005). How common such structures are within the ΛCDM model remains contentious (Yaryura, Baugh, and Angulo, 2010).

This chapter presents a percolation-based quantitative comparison between filaments found in ΛCDM models and those observed in the 2dFGRS. We rely on the percolation behaviour of the latter to define and extract filaments as described in §2.4, and apply this prescription to the 2dFGRS data in §2.5. In §2.6, we describe and present the mocks used to compare connected systems in this study, and changes made prior to applying the algorithm to these data. §2.7 details the comparison between connected systems identified in the 2dFGRS and those in the mocks, with our efforts focussing on the most luminous system in §2.8. We demonstrate in §2.9 our efforts to recover the structures characterised here using data degraded to simulated photometric redshifts. We make our concluding remarks in §2.10.

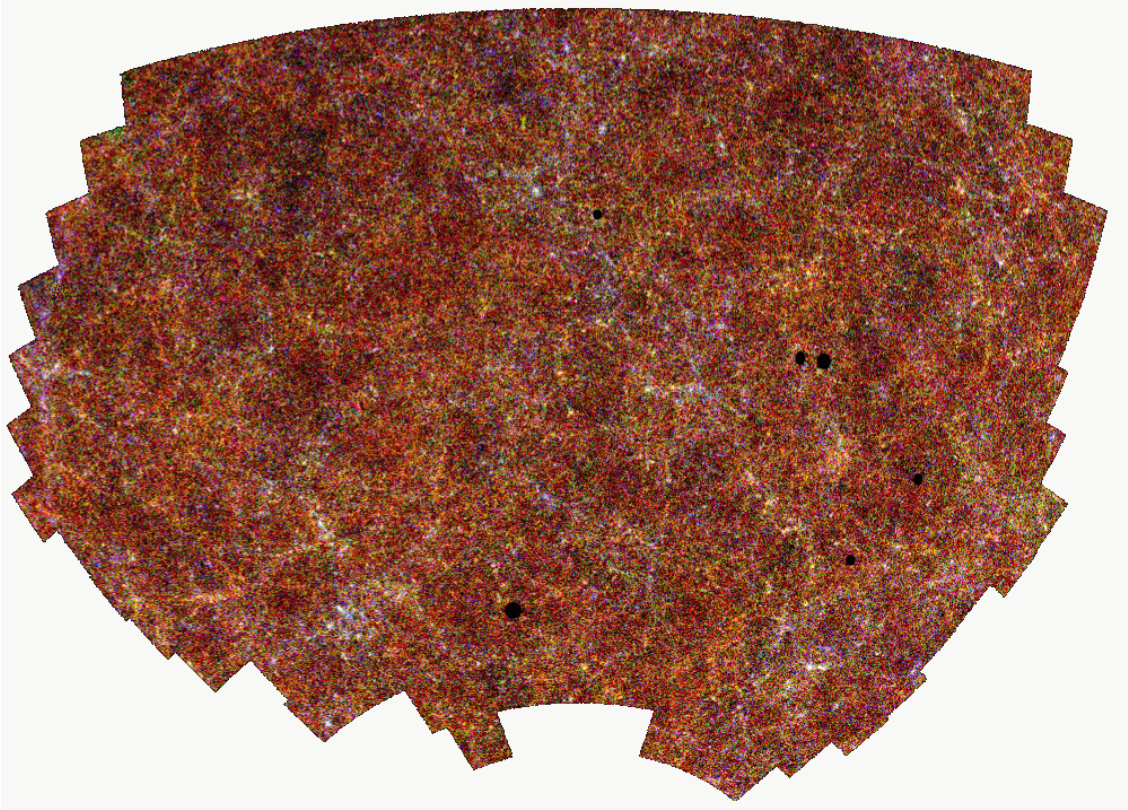

Figure 2.1: The positions of $\sim$ 2×10^6 galaxies in the APM (Maddox et al., 1990) catalogue. All galaxies between $17 \leq b_J \leq 20.5$ are shown, coded by colour such that the bright (faint) galaxies appear blue (red). The regions with high surface density are brighter than those of low surface density. (adapted from: `http://www.nottingham.ac.uk/~ppzsjm/apm`)

2.3 The Two-degree Field Galaxy Redshift Survey (2dFGRS)

This study uses the final data release of the 2dFGRS (Colless et al., 2001), a redshift survey based on two declination strips from the Automatic Plate Measuring (APM) facility survey (Maddox et al., 1990). The 2dFGRS has provided valuable (and in tandem with SDSS, complementary) insights into both cosmology and galaxy formation. For example, precision measurements of the galaxy power spectrum (Cole et al., 2005), combined with WMAP data (Spergel et al., 2007) refined estimates on the matter density parameter and the tilt of the primordial power spectrum. Measurement of the luminosity function of galaxies in the survey (Norberg et al., 2002) provided improved understanding of galaxy formation and assisted in the study of galaxy properties in a range of environments (Croton et al., 2005).

The source APM catalogue consists of digitised scans of 390 plates from the UK Schmidt Telescope (UKST) Southern Sky Survey. That survey used a $\mathrm{b_J}$ magnitude system (zero-pointed to Vega) that Blair and Gilmore (1982) connected to the Johnson-Cousins system with the colour equation:

$$\mathrm{b_J} = \mathrm{B} - 0.28(\mathrm{B} - \mathrm{V}) \tag{2.1}$$

(valid for all colours -0.1≤(B-V)≤1.6). The APM is sensitive down to $\mathrm{b_J}$ ∼22, and was used to identify the positions of galaxies sampled by the 2dF spectrograph. Figure 2.1 shows the projected positions of galaxies in the APM, where the larger-scale distribution is evident as the network of connected brighter points.

The 2dFGRS is flux-limited, with an initial faint magnitude limit set to $\mathrm{b_J} = 19.45$. This yielded a surface density of sources 30% higher than the spectroscopic fibre source density (127 fibres/square degree), and provided sufficient S/N for the faintest detectable sources within the time-frame allocated to each field. Improvements to the photometric calibrations and changes to the extinction map resulted in a position-dependent faint magnitude limit, accounted for in the survey mask, but remaining broadly $b_\mathrm{J} \lesssim 19.45$. The saturation of luminous objects on APM plates imposes a bright magnitude limit of $b_J < 14$. The 2dFGRS contains a total of 191,897 galaxies with high-quality redshifts ($\mathrm{z_{median}} = 0.11$) covering $1500\mathrm{deg}^2$ over two separate, contiguous regions towards the northern and southern galactic poles (NGP, SGP herein).

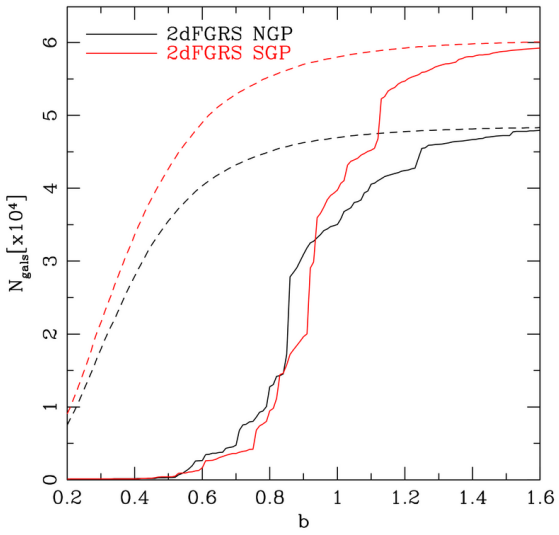

Figure 2.2: Variation in the number of 2dFGRS galaxies in structures extracted by the algorithm as the relative linking length b changes. Black lines denote galaxies from the NGP, red are galaxies from the SGP. Solid lines represent the number of galaxies in the largest system, whilst the dashed lines show the number of galaxies in all systems with at least two members.

2.4 Creating a filament detector

Our method to identify and characterise elongated 2dFGRS structures extends the concept of the Friends-of-Friends (FOF) algorithm (Huchra and Geller, 1982) used to link neighbouring galaxies together. A desirable property of an algorithm extracting large-scale structure is that there should be no preferred direction for the resultant systems. However, redshift space distortions make the line of sight a special direction in galaxy redshift surveys. The most striking consequence of non-Hubble flow velocities is to stretch galaxy clusters, creating "fingers of god" (Jackson, 1972) in the redshift-space galaxy distribution. These elongated redshift-space distortions need to be removed before searching for real structures. We achieve this by taking the "2dFGRS Percolation-Inferred Galaxy Groups" (2PIGG) catalogue of groups and clusters constructed by Eke et al. (2004a) from the 2dFGRS. These groups were identified with a FOF algorithm tested on mock versions of the 2dFGRS. A catalogue of approximately 29,000 groups out to $z \simeq 0.3$ with at least two members was constructed by linking galaxies using a cylindrical

search volume with axis ratio $l_{\parallel}/l_{\perp} = 11$ and a linking length 13% of the mean intergalactic separation. Having found galaxies belonging to groups and clusters in this way, we would like to collapse the fingers of god by placing these galaxies at their group centre positions. One complication is that Eke et al. (2004a) note that they would expect the 2PIGG catalogue to contain a few tens of per cent of interloper galaxies that are incorrectly assigned to groups. To try and correct for this inevitable mis-assignment, we choose to retain a redshift-dependent fraction

$$f(\mathrm{z}) = \frac{2 - \mathrm{z}}{2.4 + \mathrm{z}} \tag{2.2}$$

of the members assigned to each group. This expression for $f(\mathrm{z})$ is calculated from the contamination of groups found in mock 2dFGRS catalogues by Eke et al. (2004a, more details can be found in Appendix A). The randomly selected fraction $f(\mathrm{z})$ are all replaced by a single point at the group centre, whilst the remaining "interloper" galaxies are jettisoned from the list of group members and replaced to their observed redshift space positions. The first friends-of-friends pass suppresses the redshift space distortions associated with intra-group line-of-sight galaxy velocities. Note that this collapse does not account for the coherent infall of galaxies onto overdensities that will enhance and merge structure in the plane of the sky (see e.g. Kaiser, 1987; Praton et al., 1997). We then apply a friends-of-friends algorithm with a spherical linking volume to the set of remaining galaxies and group centres. The radius of this linking sphere is chosen to be b times the mean intergalactic separation at that redshift, as defined in Equation 2.7 of Eke et al. (2004a). Small linking lengths would lead to many small systems, whereas very large lengths would lead to percolation, and a single large connected component encompassing almost everything in the survey[1]. An intermediate value for b will lead to a more useful description of the structures present in the survey. This two-pass procedure provides a new and simple way to define connected structures in galaxy redshift surveys.

Figure 2.2 shows how the number of galaxies in connected structures and the number of galaxies in the largest system vary with b for both the NGP and SGP. Both the NGP and SGP regions show a rapid growth of their largest system as b increases beyond

[1]We place an upper limit of $10h^{-1}\mathrm{Mpc}$ on the physical linking length, preventing extremely large linking volumes at redshifts where the galaxy density drops rapidly.

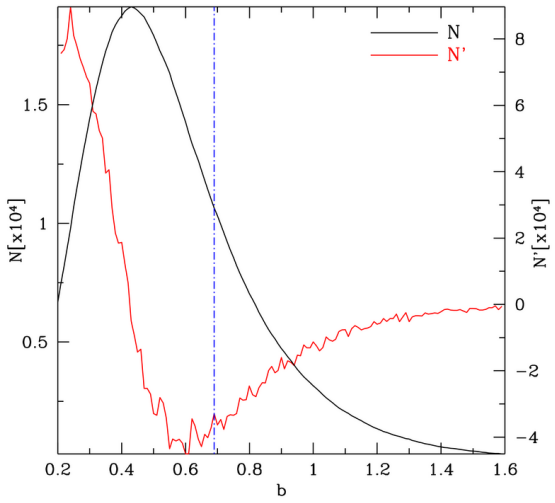

Figure 2.3: Variation in the number of structures extracted by the algorithm as the relative linking length, b, varies. The red line represents the first derivative of this function, corresponding to the rate of change of system number. We adopt a relative linking length, $b = 0.69$, close to this minimum, as denoted by the dot-dashed blue line. This corresponds to systems approximately bounded by a surface with a galaxy number overdensity of ~ 1.5.

~ 0.8. At the very largest linking lengths, not all galaxies are linked into the largest system because of the $10h^{-1}\mathrm{Mpc}$ upper limit we place on the linking length. Figure 2.3 shows the variation of the total number of connected structures and its first derivative. We pick $b = 0.69$ as a value that gives rise to an interestingly large range of system sizes. This corresponds to finding structures bounded by an irregular surface that has an overdensity of $\Delta\rho/\bar{\rho} = 3/(2\pi b^3) \approx 1.5$ (Cole and Lacey, 1996).

This choice is close to the minimum point in $\mathrm{d}N/\mathrm{d}b$, where the growth in the number of systems arising from single galaxies becoming linked matches the decrease caused by merging the structures together. The resulting systems are shown in Figure 2.4.

The abundance and extent of survey-sized connected structures will depend upon the geometry of the survey to which this algorithm is applied. Thus, while this technique is appropriate for comparing an observed data set with a mock catalogue of that particular survey, care is required when trying to infer the physical properties and abun-

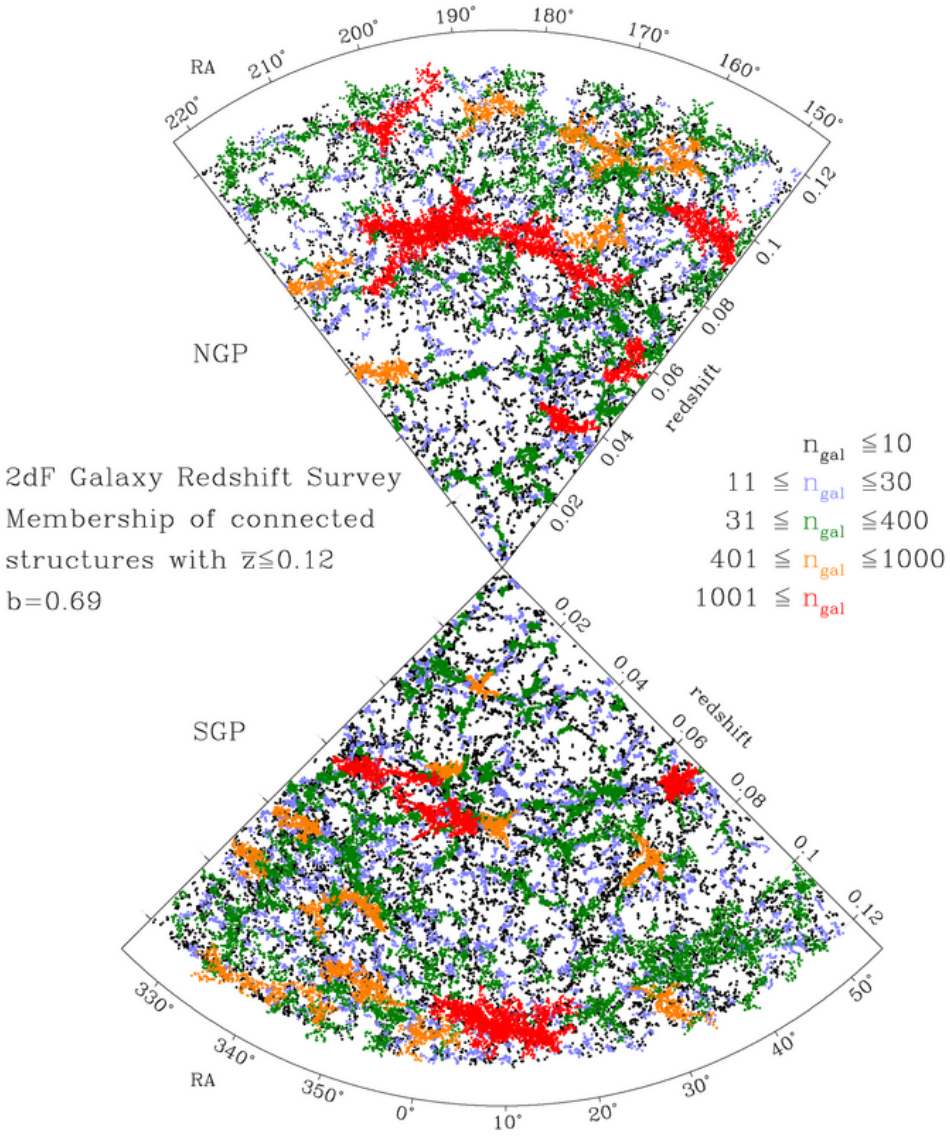

Figure 2.4: Spatial distribution of 2dFGRS galaxies in connected structures for systems with average redshift $\bar{z} \leq 0.12$ in the RA-z plane. These systems contain at least two galaxies and dot colours represent the weighted number of galaxies in the structure, where this weight takes into account the local angular incompleteness.

dance of the largest structures in the underlying distribution.

2.4.1 Connected structure luminosities

We would like to quantify the sizes of the objects found using this method in a way that (i) does not depend explicitly on the magnitude limit of the survey and (ii) assigns the

same size to a particular structure, independently of the redshift at which it is placed. Thus, rather than merely counting the number of galaxies present in each system, we define a luminosity that takes into account the flux limits of the survey. The angular variation of the flux limit in the 2dFGRS is such that it changes over the length of the elongated filamentary structures. Consequently, it is necessary to convert the observed luminosity of each galaxy to the total luminosity that would have been seen without any flux limits, rather than correcting the observed luminosity of the system as a whole. This is done assuming that the galaxy luminosity function $\Phi(L)$ is given by the Schechter function determined by Norberg et al. (2002), and using

$$L = \Sigma_i^{n_{\rm gal}} w_i L_{i,b_{\rm J}}. \tag{2.3}$$

$L_{i,b_{\rm J}}$ is calculated from the galaxy redshift the galaxy luminosity corrected for redshifting of the filter bandwidth (k-correction), evolution of the galaxy (e-correction) and contributions from galaxies below the faint (L_{min}) and above the bright luminosity (L_{max}, corresponding to $b_J < 14$) limits at that redshift. The first two of these corrections are parametrised in a similar manner to Norberg et al. (2002):

$$k + e = \frac{\mathrm{z} + 6\mathrm{z}^2}{1 + 8.9\mathrm{z}^{2.5}}. \tag{2.4}$$

At the redshift for each galaxy, one must account for galaxies which are otherwise unobservable due to the survey flux limits. Assuming the Schechter (1976) function $\Phi(L)$ describes the number of galaxies as a function of luminosity, we define the observable luminosity between L_{min} and L_{max} as L_{obs} and total luminosity as L_{tot} such that:

$$L_{obs} = \int_{L_{min}}^{L_{max}} L\Phi(L)dL, \;\; L_{tot} = \int_0^{\infty} L\Phi(L)dL, \tag{2.5}$$

$$\text{where } \; \Phi(L)dL = \phi^{\star}\left(\frac{L}{L^{\star}}\right)^{\alpha} \exp\left(-\frac{L}{L^{\star}}\right)\frac{dL}{L^{\star}}. \tag{2.6}$$

We show this in Figure 2.5, where the (redshift-dependent) shaded regions of the luminosity function highlight luminosities unobservable due to the flux limits. To compensate for the undetected flux when calculating the luminosity of the entire structure, we scale the luminosity of each galaxy accordingly. For a k+e corrected galaxy luminosity $L_{i,b_{\rm J}}^{\rm k+e}$ the total luminosity $L_{i,b_{\rm J}}$, including contributions from outside the survey

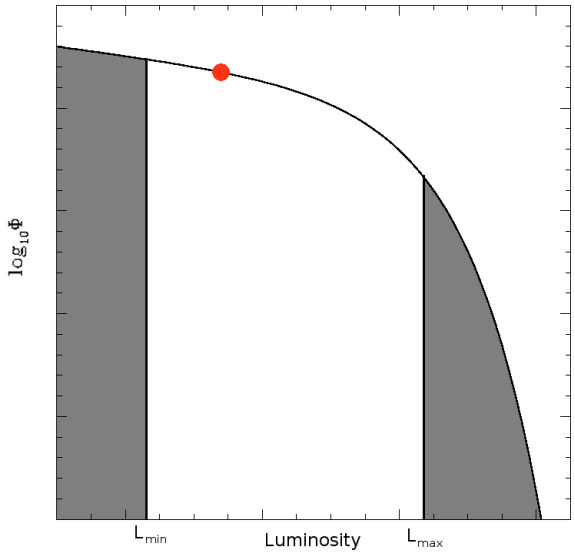

Figure 2.5: The unobserved luminosity associated with a detected galaxy (red dot) of arbitrary redshift. The shaded regions indicate the unobserved flux arising from the faint magnitude limit (left shaded region) and bright magnitude limit (right shaded region). The luminosity of the galaxy must be corrected to account for this unobserved flux.

luminosity limits is:

$$L_{i,b_{\rm J}} = L_{i,b_{\rm J}}^{\rm k+e} \frac{L_{tot}}{L_{obs}}, \tag{2.7}$$

$$= \frac{L_{i,b_{\rm J}}^{\rm k+e}}{\Gamma(\alpha+2, L_{max}/L^{\star}) - \Gamma(\alpha+2, L_{min}/L^{\star})} \tag{2.8}$$

where Γ is the incomplete Gamma function. We adopt the Schechter function parameters $(L^{\star}, \alpha) = (10^{10}\, h^{-2}\, {\rm L}_{\odot}, -1.21)$ from the (Norberg et al., 2002) study of the 2dFGRS luminosity function. The total luminosity of each connected structure is calculated by summing up all the corrected galaxy luminosities for galaxies within that system, taking into account the weighting factors that describe the local incompleteness of the survey. An alternative approach would be to correct the summed luminosity of the connected structure. However, because such systems are extended along the redshift direction, different regions will require different corrections.

Given the flux limit of the 2dFGRS, the fraction of the total luminosity that is observed drops beneath a half at redshifts exceeding z=0.12. Moreover, the linear relations used to derive Equation 2.2 begin to break down beyond this redshift. For these reasons,

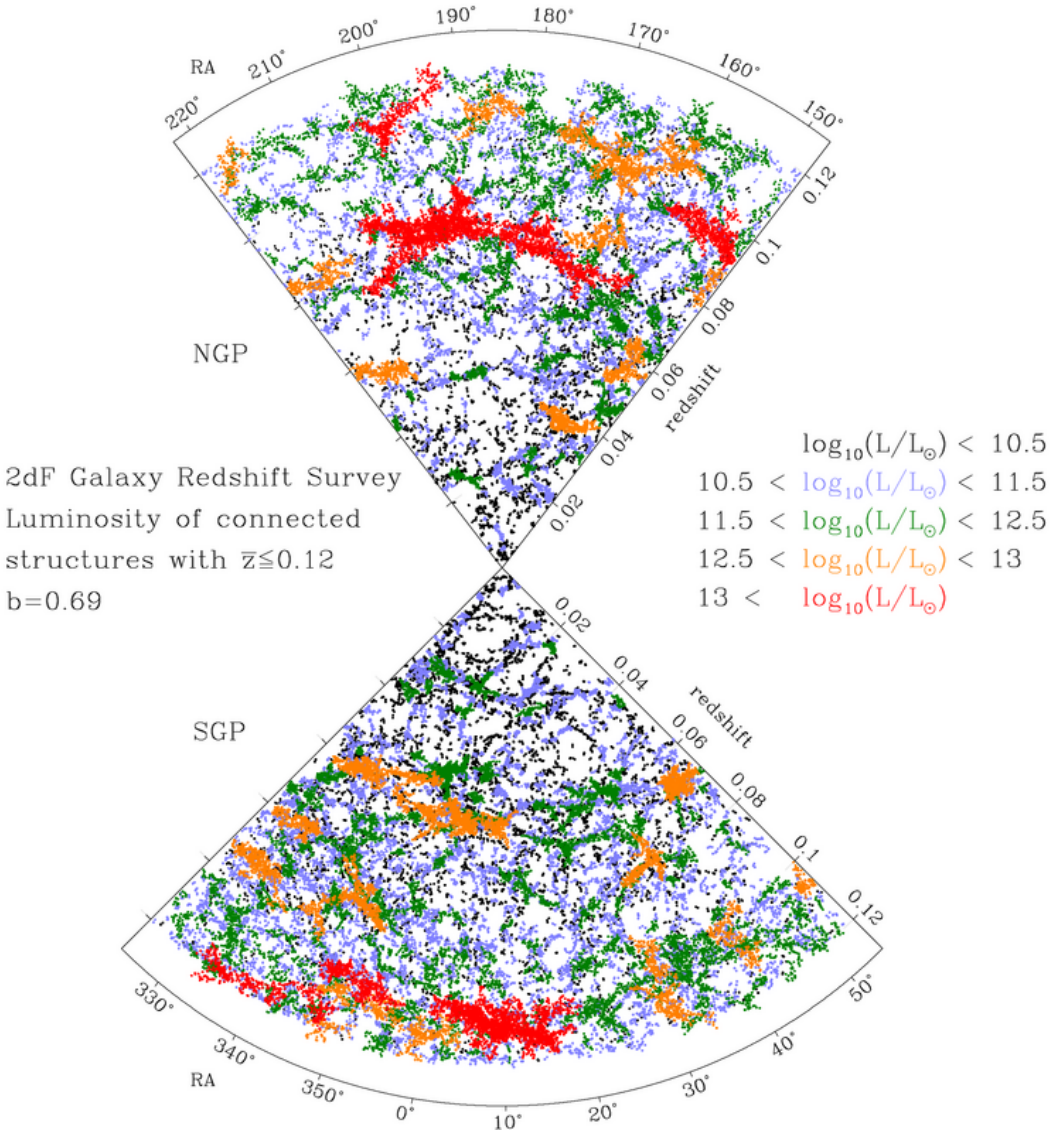

Figure 2.6: Spatial distribution of 2dFGRS galaxies in connected structures in the RA-z plane. Colours represent the total system luminosity in units of $\log_{10}(L/h^{-2}L_{\odot})$.

we will restrict our analysis to structures with $\bar{z} \leq 0.12$.

2.5 Filaments in the 2dFGRS

Figure 2.6 shows the systems found within the 2dFGRS colour coded according to their luminosity. Comparing with Figure 2.4, it is apparent how the luminosity picks out structures at larger distances more efficiently than the membership plot, which includes

only galaxies within the flux limits. A total of 95,010 galaxies are linked into 7,603 systems containing at least two members and mean redshifts no greater than 0.12. Of these, 3,018 contain only two members. Almost 87 per cent of galaxies at $\mathrm{z} \leq 0.12$ are placed into a connected structure.

One large filamentary-structure stands out in each of the NGP and SGP wedges. These systems trace out the same overdensities apparent in the 2PIGG distribution (Eke et al., 2004b), the smoothed galaxy density map (Baugh et al., 2004) and the reconstructed density field (Erdoğdu et al., 2004) of the 2dFGRS. The largest NGP object, at $\mathrm{z} \sim 0.08$, corresponds to the large RA end of the *Sloan Great Wall* highlighted by Gott et al. (2005). At a total $\mathrm{b_J}$-band luminosity of $7.8 \times 10^{13}\ h^{-2}\,\mathrm{L}_{\odot}$, this is about 20 per cent more luminous than the largest system in the SGP, which lies at $\mathrm{z} \simeq 0.11$ and RA $\simeq 10^{\circ}$. The extents in RA of these largest NGP and SGP systems in comoving coordinates are $198h^{-1}\mathrm{Mpc}$ and $99h^{-1}\mathrm{Mpc}$ respectively. While the NGP system contains twice as many members as that in the SGP, it is very nearly broken into two pieces around RA $\simeq 185^{\circ}$, where the galaxy density drops off considerably.

More locally, a continuation to lower declinations of the CfA Great Wall (Geller and Huchra, 1989) is seen at z=0.02 in the NGP, although the algorithm breaks this up into a few different components.

Some average and extreme properties of the systems identified in the 2dFGRS are listed in Table 2.1. In more detail, the correlation between the luminosity and weighted (to account for the local angular incompleteness of the survey) membership of connected structures is shown in Figure 2.7. The second largest system contains at least twice as many members as the third largest one and almost 3 times as much luminosity, making the largest NGP and SGP structures stand out from the remaining systems. The scatter around the mean relation reflects the range of redshifts in the flux-limited survey. While the object luminosity is corrected to take account of this flux limit, the weighted number of galaxies is not.

Figure 2.8 shows how connected system luminosity varies with redshift, with the lower envelope representing the total luminosity of 2 galaxies at the flux limit. The geometry of the survey precludes finding very luminous structures at low redshift because of the small volume sampled, but the greater volume available at larger redshifts is sufficient to contain larger, more luminous filamentary structures. The largest NGP and SGP

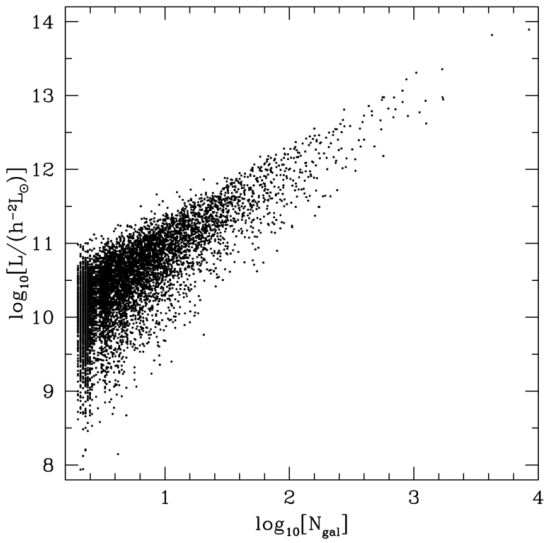

Figure 2.7: The relation between object luminosity, L, and N_{gal}, the weighted number of galaxies in systems with $\bar{z} \leq 0.12$. The two large superstructures are noticeable here in the extreme upper right of the plot.

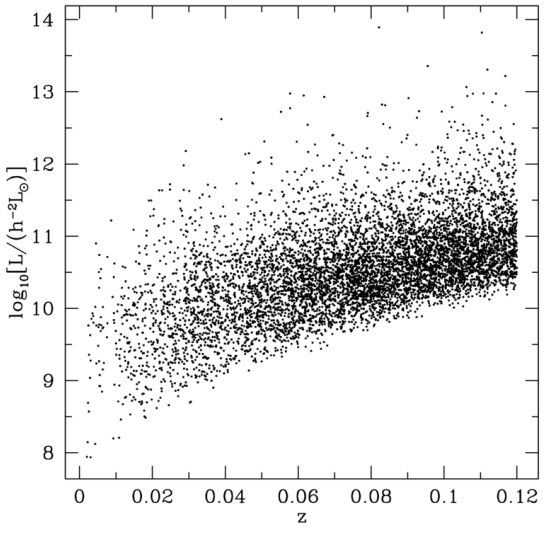

Figure 2.8: The distribution of system luminosities with increasing redshift. The 2dFGRS flux limit is evident here as the gradient in minimum luminosity.

Survey	N	f (%)	$\log_{10}(N_{\rm gal})$	$\log_{10}\bar{L}$	$\log_{10} L_{max}$	l_{max}
2dFGRS	7603	87.7	5.06	11.16	13.89	198
2BASICC	8023 ± 250	85.9 ± 0.8	5.07 ± 0.04	11.08 ± 0.03	$13.44^{+0.15}_{-0.23}$	81 ± 19
HV	8253 ± 135	82.0 ± 0.8	5.11 ± 0.02	11.10 ± 0.03	$13.55^{+0.14}_{-0.20}$	93 ± 27

Table 2.1: Properties of connected structures with $\bar{\rm z} \leq 0.12$ identified in the 2dFGRS and mock surveys. Mock values are the mean over all 50 2BASICC and 22 HV mock surveys with the uncertainties being the standard deviation of individual surveys from these mean values. N is the total number of connected systems within the catalogue, f the fraction of galaxies out to ${\rm z} = 0.12$ in systems and $N_{\rm gal}$ is the total number of galaxies out to the same redshift. The fifth and sixth columns describe the average and maximum object luminosities in units $\log_{10}(L/h^{-2}L_{\odot})$. We give the comoving scale l_{max} (in $h^{-1}{\rm Mpc}$) of the largest structure identified in the survey in the final column. This is defined as the largest in extent of galaxy members in the redshift, RA or dec directions.

systems are once again conspicuous at the top of the figure.

Although comparisons could be made with similar 2dFGRS studies such as Pimbblet et al. (2004), we reiterate here our final comments in §2.4. Our choice of parameters adopted for the filament-finding part of the algorithm, and the way we define connected structures is both subjective and specific to the survey geometry (and source galaxy catalogue). It is therefore non-trivial to directly compare our findings to those of alternative studies without a proper appraisal of how such structures are defined and selected. It is in this respect that we turn to our comparison with mock 2dFGRS surveys to draw conclusions from our results.

In the Pimbblet et al. (2004) analysis, the source cluster catalogue (De Propris et al., 2002) was compiled from a variety of different sources: the two Abell catalogues (Abell, 1958; Abell et al., 1989), the Edinburgh-Durham Cluster Catalogue (EDCC; Lumsden et al., 1992) and the APM Cluster Catalogue (Dalton et al., 1997). Through the introduction of clusters from many different sources, it becomes difficult to compare the selection

function[2] of the resultant merged catalogue, and consequently any filaments identified from them. The Pimbblet et al. (2004) study selects filaments based on cluster pairs of up to 10 degrees separation with similar redshifts. These pairs are visually inspected for connecting chains of galaxies, and classified based on a morphological scheme defined in (Colberg et al., 2005).

By contrast, the approach we present here includes groups derived from only one source, with a well-defined selection function (Eke et al., 2004a) and an optimally-tuned group-finder motivated by studies from realistic mock surveys. We automate the process of detecting large-scale structures in this dataset to prevent any bias arising from human-error. Differences in selection criteria mean the number of detected systems varies significantly between the two studies (7,603 vs. 805). To better evaluate our findings in the 2dFGRS data, we turn to mock surveys.

2.6 Mock Surveys

One of the most successful aspects of the current cosmological model is the agreement between observational data and computer-generated universes. Data from large surveys such as the 2dFGRS have led to improved constraints on the cosmological and galaxy formation models used, in turn, to produce even more realistic simulations and mock surveys. To assess the significance of structures within the 2dFGRS we perform the procedure described above on a series of mock surveys; this section describes the mocks used in our analysis. In §2.6.1 we describe the surveys already generated for the 2dFGRS, and in §2.6.2 we detail the creation of a new set of realistic mocks.

2.6.1 The Hubble Volume mock 2dFGRS surveys

The Hubble Simulation (Evrard et al., 2002), created by the Virgo Consortium[3], was a series of simulations with different cosmologies. In this study, we use data derived from a model with Ω_M=Ω_{DM}+Ω_b=0.3, h=0.7, Ω_Λ=0.7, and σ_8=0.9. This particular simulation contained 10^9 particles over a volume of $27h^{-3}\mathrm{Gpc}^3$.

[2]Including, for example, well-known projection and completeness issues in the Abell catalogues - see §1.4.2.

[3]`http://www.virgo.dur.ac.uk`

Cole et al. (2005) describe how galaxies were assigned to dark matter particles. A continuous density distribution was first estimated by smoothing the particles with a $2h^{-1}\mathrm{Mpc}$-width Gaussian kernel. Galaxies were assigned to particles exceeding a density-dependent probability threshold that was defined in order to reproduce the 2dFGRS clustering signal. Twenty-two Hubble Volume (HV) surveys were extracted from the simulation, with observer positions selected to resemble conditions in the Local Group (and in particular a match to the observed-frame velocity and direction of the CMB dipole). Norberg et al. (2002) assign $\mathrm{b_J}$-band magnitudes to galaxies, based on their redshift by sampling from the 2dFGRS Schechter function. This particular prescription did not include descriptions of luminosity or colour-dependent clustering.

2.6.2 Building new mocks: The 2BASICC surveys

In this section we describe the creation of a new set of 2dFGRS mock surveys complementary to the HV surveys. We first describe the dark matter simulation that the galaxies are applied to. We next discuss the semi-analytic approach that generated the population of galaxies. Finally, we take the supplied galaxies (C. Baugh, priv. communication) generated by this prescription and extract subsamples with the correct 2dFGRS geometries and selection function.

- The BASICC simulation

We use the "Baryon-Acoustic-Oscillations at the ICC" (BASICC; Angulo et al., 2008) simulation to construct a series of 50 mock surveys dubbed 2dFGRS-BASICC (2BASICC). The BASICC simulation was created with a cosmology of Ω_M=Ω_{DM}+Ω_b=0.25, Ω_b=0.045, h=0.73, Ω_Λ=0.75, with σ_8=0.9 and spans a volume of $2.4h^{-3}\mathrm{Gpc}^3$ with 3×10^9 dark matter particles (with masses $M_\mathrm{p}=5.49\times10^{10}h^{-1}M_\odot$) evolving from z=63 to z=0. Haloes were defined as groups of at least 10 dark matter particles, identified by a friends-of-friends algorithm with an optimally chosen relative linking length $b=0.2$ (see Davis et al., 1985; Jenkins et al., 2001). This results in haloes with a minimum mass of $M_\mathrm{H}=5.49\times10^{11}h^{-1}M_\odot$. In the z=0 snapshot we study here, there are a total of 17,258,579 haloes. Simulations generally require some compromise between particle mass and simulation volume. The large simulation volume here necessitates a high

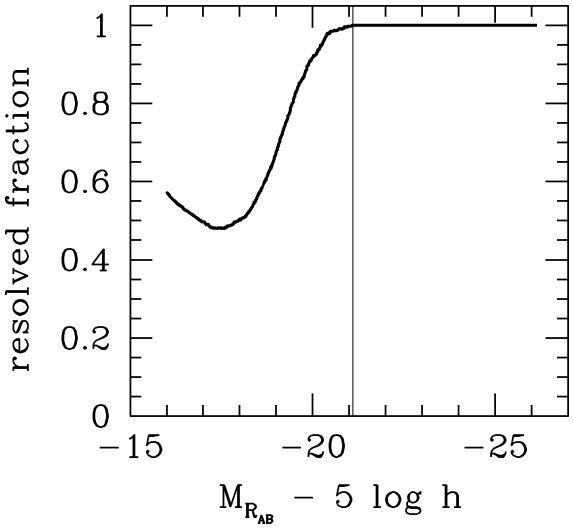

Figure 2.9: (adapted from Angulo et al., 2008). The fraction of galaxies residing in resolved haloes for the R-band. The completeness drops below one at $M_{R_{AB}} = -21.1$, requiring the addition of an "unresolved" galaxy population below this threshold.

minimum halo mass. By comparison, the Millennium Simulation (Springel et al., 2005) is some twenty times less massive in volume, but uses dark matter particles sixty-four times smaller. Moreover, the BASICC simulation features only 12 redshift snapshots, as opposed to 63 in the Millennium run.

To populate BASICC with galaxies the GALFORM (Cole et al., 2000) prescription was used to assign galaxies to dark matter haloes using the Baugh et al. (2005) model. Ideally, one makes use of the detailed halo merger trees when assigning the properties of the galaxies in each halo. However with only 12 redshift snapshots, a detailed merger history for each halo is not possible. The alternative adopted here was to define Monte-Carlo (MC) merger histories over the range of halo masses detected in the simulation. Although haloes of the same final mass undergo different merger histories, the properties of the trees (and hence through GALFORM their galaxy properties) are statistically similar enough for any halo assembly bias to be ignored (although see Gao et al., 2005; Gao and White, 2007, for a better appraisal of how assembly history impacts halo clustering). Each BASICC halo was matched to GALFORM outputs of MC merger-trees in the same mass range. GALFORM places central galaxies at the minimum potential of each halo. Lower-luminosity satellite galaxies are assigned randomly to the dark matter

member particles.

The range of central and satellite galaxy luminosities assigned to haloes in the simulation results in a gradual, rather than sharp cutoff in the completeness of the galaxy catalogue at the minimum halo mass. This can be seen in Figure 2.9, adapted from Angulo et al. 2008, which shows the completeness of mock galaxies in the R-band. The vertical line indicates the galaxy luminosity associated with the halo mass limit.

To include fainter galaxies that would have been hosted in haloes below the simulation mass resolution, GALFORM galaxies based on the Baugh et al. (2005) model were generated for masses below $M_{\rm H} = 5.49 \times 10^{11} h^{-1} M_{\odot}$ and assigned to dark matter particles not identified as halo members by the halo groupfinder.

- Creating the 2dFGRS mock surveys

In our study, we use the data described in the previous section to construct mock 2dFGRS surveys. We first calculate the $\rm b_J$ and $\rm r_f$ luminosities from the SDSS bands generated by GALFORM. For this we use the relation derived by Norberg et al. (2002):

$$\rm b_J = g + 0.155 + 0.152(g - r) \tag{2.9}$$

$$\rm r_f = r - 0.3 \tag{2.10}$$

A larger sample volume is constructed from the galaxy data by surrounding this simulation cube with 26 (i.e. 3^3-1) periodic replications. For each of the 50 mock surveys we produce, an observer position is randomly chosen within the central cube and a random direction is adopted. Both the positions and volumes of the 50 surveys are not independent: each 2dFGRS mock survey (NGP and SGP combined) occupies 0.7% of the simulation volume suggesting it is not impossible for separate surveys to share portions of the simulation[4]. The periodic replications allow observer positions close to the edge of the cube to select galaxies beyond the boundary. The direction chosen for the mock defines the orientation of the NGP and SGP survey slices, and for each of the 50 surveys we search through the cube and periodic replications to identify galaxies with $M_{\rm bJ} - 5log_{10}h \leq$-15 inside these two volumes. We exclude galaxies too distant for our connected structure analysis by applying a redshift cut of $\rm z \leq 0.13$.

[4]However, the randomised directions and application of angular and radial selection functions will limit any repetition of structure.

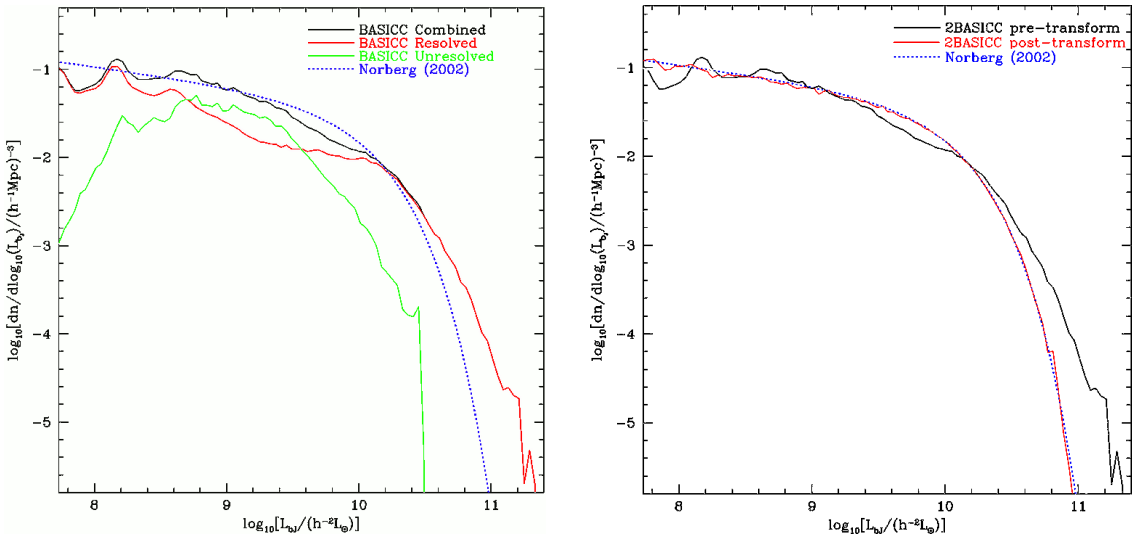

Figure 2.10: (*left*) The BASICC luminosity function. The black line represents the combined luminosity function comprising of galaxies in resolved (red) and unresolved (green) haloes. For comparison, the 2dFGRS b_J-band Schechter function (Norberg et al., 2002) is also plotted. (*right*) The luminosity functions of mock galaxies inside a 2dFGRS-sized volume. The black (red) function shows the luminosity function before (after) translation to the reference 2dFGRS-fit Schechter function (shown as a blue dotted line in both panels).

In the *left* panel of Figure 2.10 we plot the luminosity function of galaxies within the full simulation cube, split between those in resolved (*red*) and unresolved (*green*) haloes. The composite luminosity function, shown in *black*, benefits from the inclusion of the unresolved population faint-wards of L* but over-estimates the number of bright (L$\sim 10^{11}\, h^{-2}\, \mathrm{L}_\odot$) galaxies. This overabundance is often corrected by including AGN feedback into the semi-analytic model (see Bower et al., 2006). Because the sub-L* population is also inconsistent (although less so) with the Norberg et al. (2002) 2dFGRS Schechter fit (*dotted blue*; $(L^\star, \alpha, \phi^\star) = (10^{10}\, h^{-2}\, \mathrm{L}_\odot, -1.21, 1.61 \times 10^{-2})$), we calculate the shifts required in bins of luminosity in order to improve this fit. These adjustments, calculated from the full simulation cube, are applied to each of our 50 sample volumes[5]. The *right* panel of Figure 2.10 shows the effect of this transformation on the luminosity function

[5]We adopt this approach, rather than calculating the luminosity shifts in each of the 50 survey volumes, in order to maintain some degree of cosmic variance within the set of surveys.

of a 2dFGRS mock volume. The black line again shows the original luminosity function, whilst the red line indicates the transformed function, compared to the Norberg et al. (2002) 2dFGRS Schechter fit (blue dotted line).

Each galaxy in the survey volume is assigned a redshift. The recession velocity v_{rec} of each galaxy arises from both the Hubble flow (Equation 1.1) and the peculiar velocity. This latter component may come from the bulk or cluster-driven motion of galaxies with respect to the Hubble flow:

$$\mathrm{z_{gal}} = \mathrm{z_H} + \frac{\underline{\mathrm{v}} \cdot \hat{\underline{\mathrm{r}}}}{\mathrm{c}} \tag{2.11}$$

where $\mathrm{z_H}$ is the cosmological redshift and $\underline{v} \cdot \hat{\underline{r}}$ is the radial component of the galaxy's velocity. Velocity 3-vectors (supplied by `GALFORM` from the dark matter velocity field) are used to calculate the line-of-sight velocity components.

Finally we apply the angular selection masks and radial selection function. These enforce the 2dFGRS position-dependent flux limits, assign galaxy weights (to account for unselected galaxies) to the mock galaxies, and produce surveys ready for our analysis[6]. Figure 2.11 shows an example of one such flux-limited survey, with a magnitude cut applied to illustrate the features of the volume.

2.6.3 Modifying the filament detector for use with mock surveys

Previous studies have shown semi-analytic models tend to place too many low-luminosity galaxies into galaxy clusters (Eke et al., 2004b; Gilbank and Balogh, 2008; Kim et al., 2009). This can be seen in Figure 2.12 taken from Eke et al. (2004b), where we plot the luminosity function of galaxies in groups of mass $10^{14}h^{-1}M_{\odot}$. There is a notable excess of $\simeq 10^9\ h^{-2}\,\mathrm{L}_{\odot}$-luminosity galaxies in the model data. Because the mock luminosity function has been forced to match that of the observations, this implies the model will lack low-luminosity galaxies in lower density regions. This known problem will affect the structure finder.

To reduce the impact of this difference between the model and observations, we allow ourselves the freedom to jettison a smaller fraction of galaxies from the model groups than given by Equation 2.2 for the real 2dFGRS. This decreases, in the vicinity of the

[6] All 50 mock surveys are available online at `http://astro.dur.ac.uk/~dmurphy/2basicc`

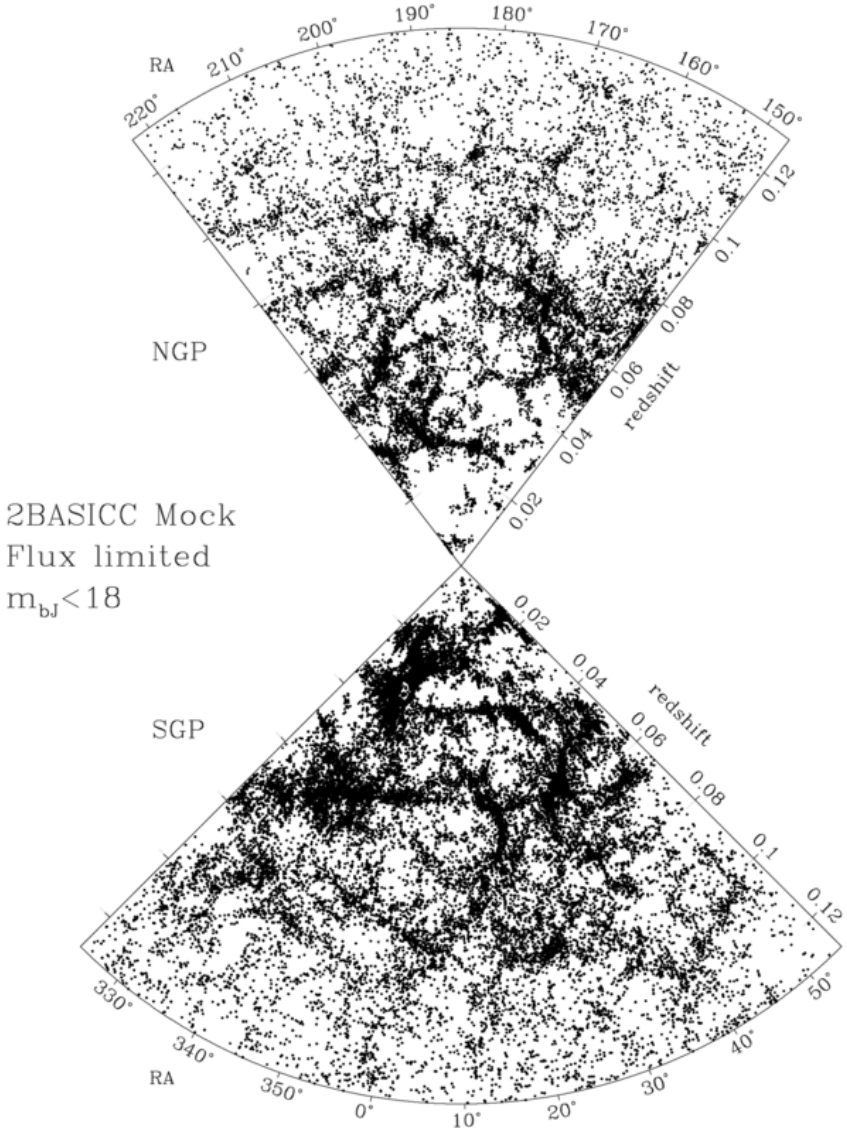

Figure 2.11: A flux-limited 2BASICC mock used in the filament-finder analysis. Both the radial selection function and the angular mask have been applied. For clarity, this plot shows galaxies with magnitudes of $b_J < 18$.

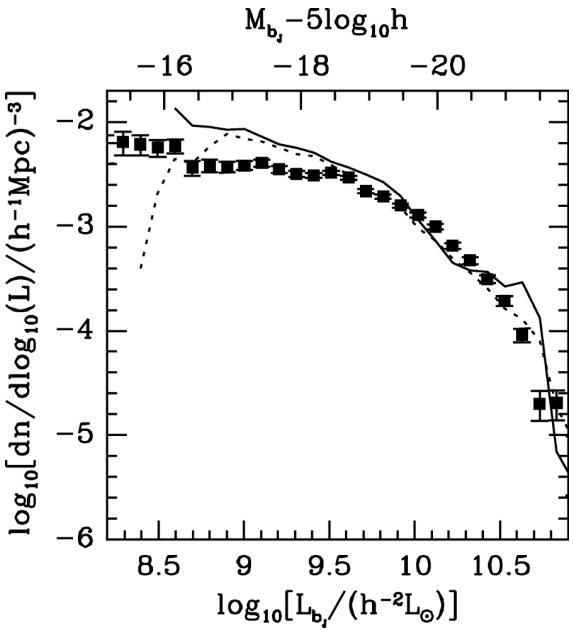

Figure 2.12: (adapted from Eke et al., 2004b) The luminosity function of galaxies in groups of mass $10^{14}h^{-1}M_{\odot}$. The 2dFGRS groups (filled squares) are compared to the equivalent luminosity function recovered from mock surveys (dotted line). For reference, the input model group luminosity function (solid line) is included. Group masses from 2dFGRS data are dynamical mass estimates from member velocity dispersions. There is a clear excess of mock galaxies at $\simeq 10^{9}\,h^{-2}\,\mathrm{L}_{\odot}$.

groups, the number density of points used for the structure-finding sweep of the friends-of-friends algorithm to an amount similar to that in the real survey. We achieve this in the model by multiplying $f(\mathrm{z})$, as given by Equation 2.2, by a constant, χ. Where $\chi > 1$, this implies a higher fraction of galaxies remain in the groups.

In order to determine an appropriate value for χ, we have measured the distribution of system orientations defined as

$$\theta = \tan^{-1}\left(\frac{\Delta l_{\mathrm{z}}}{\Delta l_{\phi}}\right) \tag{2.12}$$

where Δl_{z} represents the range of the member galaxies in the redshift direction and Δl_{ϕ} is the larger of the ranges of member galaxies in the RA and dec directions. Thus, $\theta = \pi/2$ for a radial object and 0 for one lying perpendicular to this. We use the greater of Δl_{z} and Δl_{ϕ} to describe the scale size of the connected structure.

Figure 2.13 shows the cumulative probability distributions of system orientations

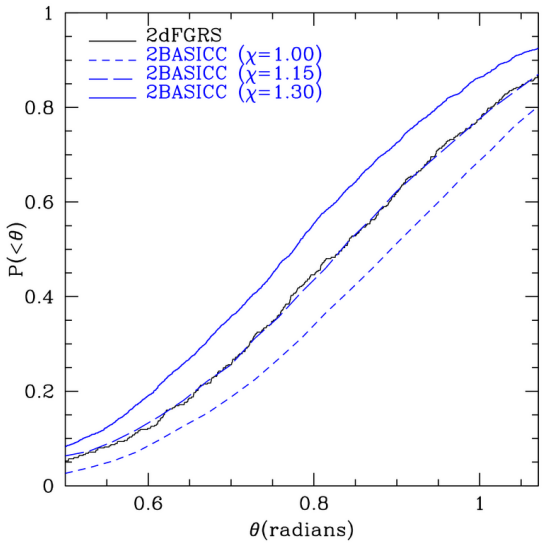

Figure 2.13: Cumulative probability distributions of connected system orientations for all objects containing at least 20 galaxies. Results are shown for the 2dFGRS and for averages of 10 2BASICC mocks. The mock distributions are derived from three different choices of $\chi = 1.0, 1.15$ and 1.30, as indicated in the legend.

for structures containing at least 20 galaxies in the 2dFGRS and those recovered from 10 mock surveys using three different values of χ[7]. We adopt a membership cut of 20 galaxies to reduce the inclusion of poor systems that, with their less constrained orientations, may add random noise to the distribution. It is apparent, when treated in the same way as the real data (i.e. $\chi = 1$), that the model contains too many objects aligned along the line of sight. This is a result of too many low luminosity galaxies being placed into the redshift space volumes occupied by the model groups. When the "interloping" galaxies are jettisoned from the groups found in the mock catalogues, enough of these additional low luminosity galaxies are placed along radial lines that they bias the orientation distribution. Increasing χ retains a higher fraction of the initially grouped galaxies in the groups, reducing the number of interlopers returned to the field, and decreasing the number of radially aligned objects found in the second pass of the friends-of-friends

[7]One might imagine that randomly oriented connected structures would be uniformly distributed with θ. However, since systems often contain more than two galaxies, which are generally not colinear, the definition of θ leads to connected structures preferentially avoiding values towards the ends of the range $[0, \pi/2]$.

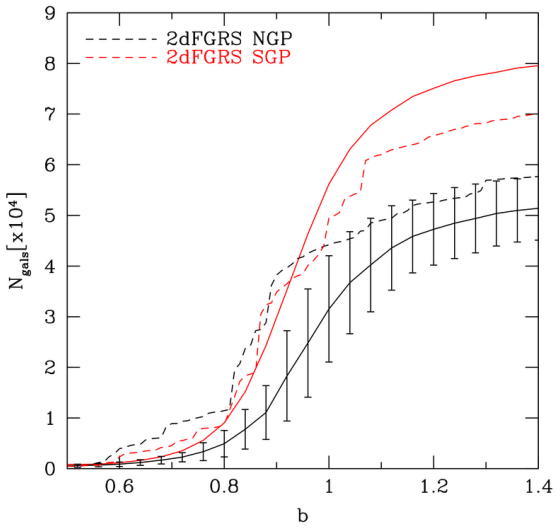

Figure 2.14: The weighted number of galaxies in the largest structure for NGP (black) and SGP (red) systems with $\bar{z} \leq 0.12$, including any members with redshifts greater than this limit subject to the mean system redshift remaining below it. In both cases the dashed lines show the 2dFGRS data, whilst solid lines represent the mean number of galaxies in the largest object across 50 mock surveys. In the NGP case, we include also error bars representing the standard deviation of these surveys around the mean.

algorithm. A value of $\chi = 1.15$ produces a mock orientation distribution that is, according to a Kolmogorov-Smirnov test, indistinguishable from that found in the 2dFGRS. This is chosen as the default value for these 2BASICC mocks throughout this chapter. This procedure was also applied to the set of 22 HV mocks, and a value of $\chi = 1.11$ was derived to match the 2dFGRS filament orientation distribution.

We follow almost the same procedure as described in §2.4 to define connected structures in the mocks, with the only difference being that the fraction of galaxies retained in the groups, $f(z)$ in Equation 2.2, is increased by a factor χ, as described above. The impact of this choice on the results is discussed in the following section.

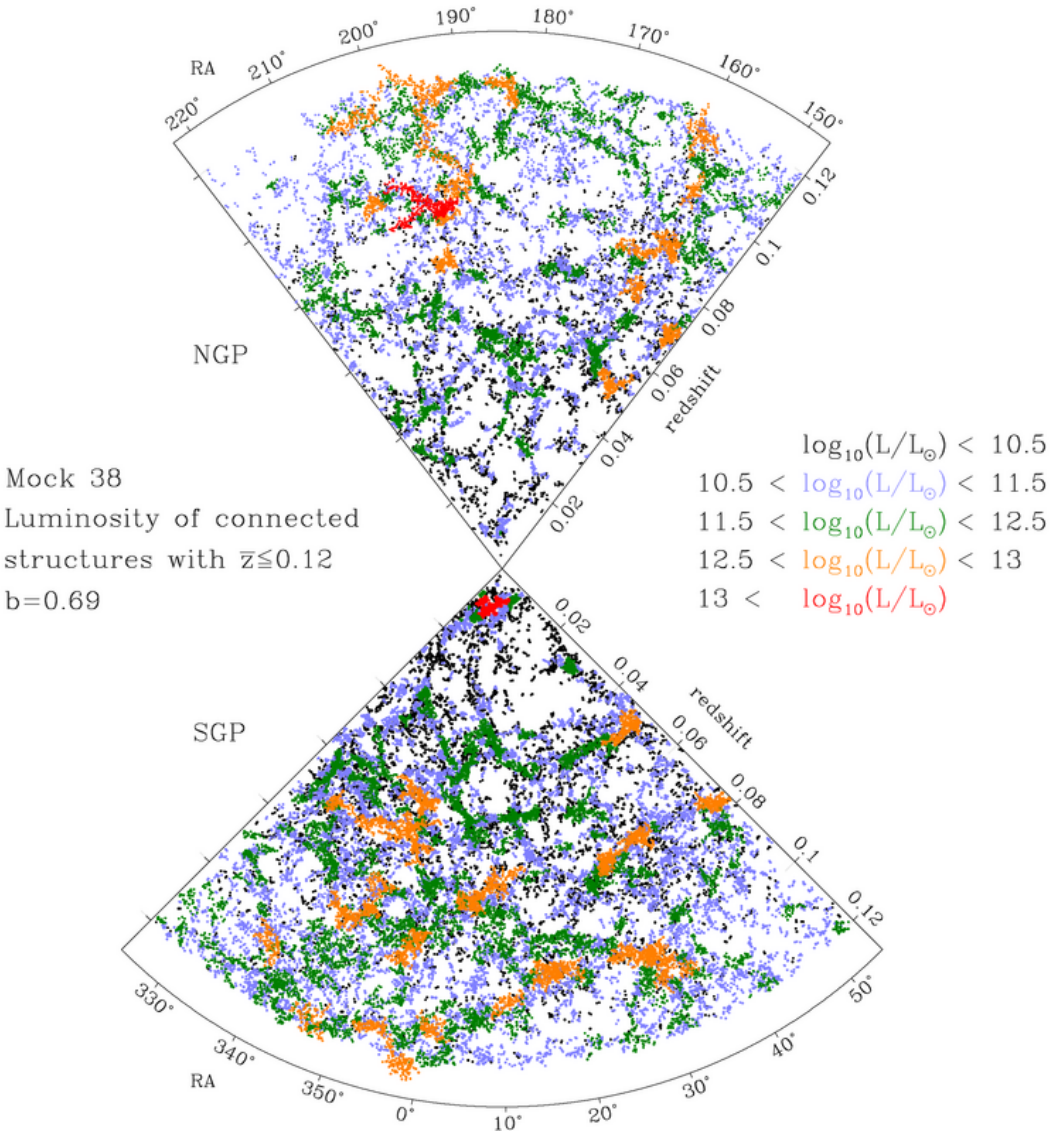

Figure 2.15: Spatial distribution of mock galaxies in connected structures in the RA-z plane for 2BASICC mock #38. This survey is notable for the large structure identified close to the observer in the SGP. This system is a 4-σ outlier in galaxy membership across the 50 mock surveys.

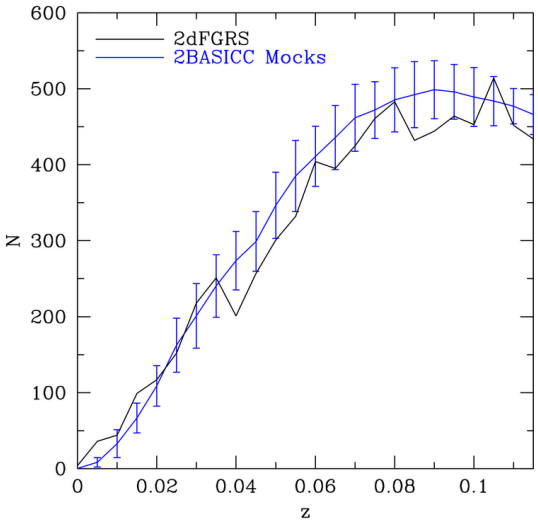

Figure 2.16: The redshift distribution of all objects with at least two members. The blue line represents the mean number of connected structures as a function of redshift across 50 mock surveys, with error bars representing their standard deviation around the mean. The black line corresponds to systems detected in the 2dFGRS.

2.7 Comparing filament catalogues

Figure 2.14 shows how the number of galaxies in the most populated system grows as b is increased for each survey region. While the behaviour is broadly similar to that of the largest filament in the 2dFGRS, the onset of percolation in mock catalogues is delayed by about 0.1 times the mean intergalaxy separation. As a consequence, at $b = 0.69$, the largest 2dFGRS system (located in the NGP) is a significant outlier, being more populated than the corresponding structure in all but one of the 50 mock catalogues. The mock system containing more galaxies than the largest one in the 2dFGRS is placed at $\mathrm{z} \sim 0.01$, is comparatively less luminous, and results from the randomly chosen observer being put very near to a large galaxy cluster. This survey (from 2BASICC mock #38) can be seen in Figure 2.15. The structure in question (in the SGP) has 7600 galaxies, but is $\sim$25% fainter than the brightest 2dFGRS system.

The larger value of χ adopted for mock surveys means that the number of galaxies and group centres used for the algorithm is on average ~ 1.3 per cent lower than in the $\chi = 1$ case. Even if we adopt a linking length scaled by the inverse cube root of mock

centre numbers to 2dF centre numbers (corresponding to $b = 0.72$), this does not significantly affect the discrepancy between the number of galaxies in the most populated systems in the mock or real 2dFGRS.

The redshift distribution of the structures is shown in Figure 2.16. At $\mathrm{z} \lesssim 0.025$, where mocks are known to be missing low luminosity galaxies, the mock surveys contain fewer objects than the real 2dFGRS. The main reason for this is actually not the incompleteness in the mocks, but the fact that too many low luminosity galaxies are placed into large groups, reducing the number available to form other small systems. This local volume represents only a small fraction of the survey.

For redshifts greater than 0.04, the number of real 2dFGRS structures is typically slightly below the mean of the 50 mock surveys. This is reflected in the first column of Table 2.1, which shows that the total number of systems in the 2dFGRS is $(1-2)\sigma$ beneath that of the mocks, despite the fact that a slightly higher fraction of galaxies are placed into the 2dFGRS structures. The total number of galaxies in the 2dFGRS and mock surveys matches well by construction, but the excess of mock galaxies placed into groups means that fewer are available in lower density regions for linking together small systems. The relatively high fraction of galaxies placed into structures and low number of structures in the 2dFGRS leads to a larger mean luminosity. This difference can be removed by not including the two most luminous 2dFGRS systems in the calculation.

The distribution of system luminosities is shown in the *left*-hand panel of Figure 2.17. The mocks have a relative lack of structures at $L \lesssim 10^9\, h^{-2}\, \mathrm{L}_\odot$, and more than the real 2dFGRS at $L \simeq 3 \times 10^{10}\, h^{-2}\, \mathrm{L}_\odot$ (which is the peak of the distribution, corresponding to approximately to two $L^\star$ galaxies), and a paucity of filamentary systems like the largest ones in the 2dFGRS. As stated above, the difference between the model and real distributions at low luminosities arises mostly because the lowest luminosity galaxies in the model are more likely to be placed into larger groups and hence are not available to form very low luminosity systems. The deficit of lower luminosity galaxies outside groups impacts in two opposing ways upon the most luminous model structures. They tend to gain luminosity because their groups are slightly more luminous than those of corresponding mass in the 2dFGRS. However, the lack of low luminosity galaxies in the lower density regions makes it less likely that large structures will join together. It is this second effect that is more important, resulting in none of the 50 mock surveys producing

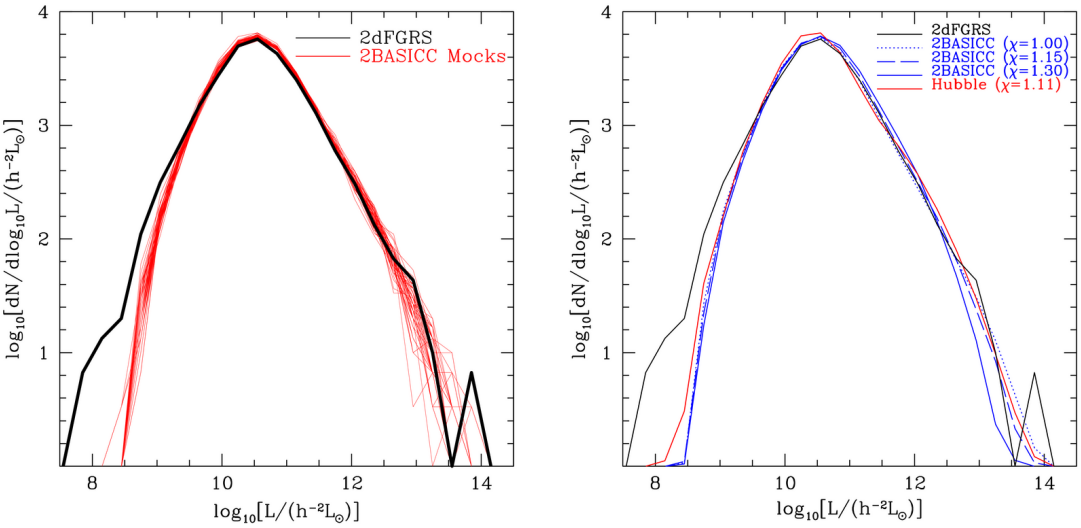

Figure 2.17: (*left*) The distribution of luminosities for all structures with a minimum membership of two, out to a redshift $z = 0.12$. The red lines show the distribution for each of the 50 2BASICC mock catalogues. The black line shows the distribution for structures in the 2dFGRS. (*right*) Structure luminosity distributions for different values of χ in these surveys and for the HV catalogues. Mock survey distributions have been averaged over the 50 surveys in the 2BASICC and the 22 in the HV simulations.

a pair of systems as luminous as the most luminous 2dFGRS pair.

Given that many differences between the real and mock filament luminosity distributions result from the different spatial distributions of low luminosity galaxies in the real and mock surveys, and that we have used a different value of χ for real and mock surveys, one might reasonably ask what changes when $\chi = 1$ is used for the mocks. This is shown in the *right*-hand panel of Figure 2.17, where three different χ values are used for the 2BASICC mocks. Increasing χ retains more galaxies in the groups, leaving fewer galaxies to help the filament finder link together larger structures. This leads to a decrease in the luminosity of the most luminous filaments. Decreasing χ leads to an increase in the luminosity of the most luminous filaments, but even for $\chi = 1$ there are still no surveys with two filaments at least as luminous as the second most luminous 2dFGRS filament. Nevertheless, we do obtain two surveys with a filament more luminous than the brightest 2dFGRS filament. However, as shown in Figure 2.13, this comes at the

expense of producing a set of filaments that are significantly more radially oriented than those found in the 2dFGRS.

Also shown in this panel is the distribution of filaments found in the 22 HV mock catalogues of Cole et al. (2000). The bottom row of Table 2.1 contains statistics for the connected structures found in these HV mocks. These surveys were constructed with both a different cosmology (primarily a smaller dark energy component) and different methods for assigning galaxies to dark matter particles. The connected structure luminosity distributions are broadly similar to the 2BASICC, with the abundance of the most luminous filaments being almost unchanged. Again, we find none of the surveys investigated contain a pair of filaments as luminous as the two detected in the 2dFGRS. In a study of the three-point correlation function of 2dFGRS galaxies, Gaztañaga et al. (2005) state they fail to detect 2dFGRS superstructure-scale systems in their 22 mock HV surveys, suggesting the probability a mock survey will feature such systems is $<1/22$ ($\sim$5%). In our study, we can revise this upper limit on the probability of finding such structures in mock 2dFGRS surveys down to 1.4%. There is a slight increase in the faintest systems identified in the HV Mocks (the *red* line) relative to 2BASICC (*blue* lines). As this luminosity probes the environment local to the observer, differences may potentially arise for two reasons. First is the way the observer is placed in the simulation. Although entirely random for 2BASICC, HV observers are constrained to the kinematics of the Local Group. This will permit 2BASICC observers to be placed in a more diverse range of environments, including voids bereft of local structure. Alternatively, this slight difference could be accounted for by the lack of luminosity-dependent clustering in the HV mocks. In the 2BASICC surveys, the clustering amplitude will be lower for fainter galaxies, resulting in fewer low-luminosity connected structures. Overall however, the difference in luminosity distribution between these two mock survey approaches are negligible.

The distribution similarities arise from the selection of different χ values, suggesting such a parameter could in principle be used as a probe of cosmology and galaxy formation model. However, with only a marginal difference between the HV luminosity distribution and those of 2BASICC in Figure 2.17 (save for the $\chi_{\rm 2BASICC}$=1.30 case), the use of an alternative, more sensitive, parameter may be preferred. Nevertheless, surveys probing to higher redshifts would permit the study of how, under differing cosmologies,

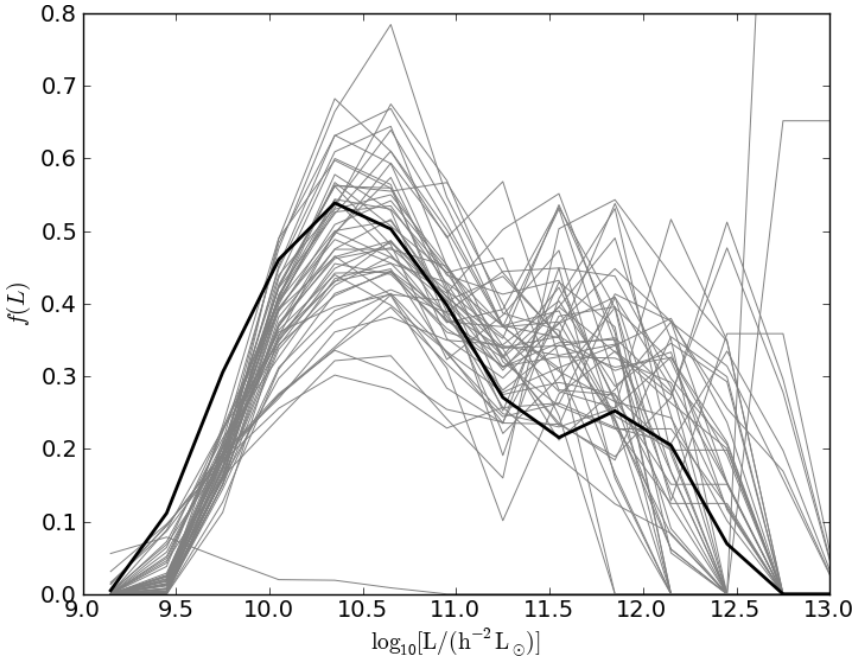

Figure 2.18: The normalised luminosity contributions from galaxies making up the most luminous filament in each mock survey (grey) and the 2dFGRS (black). Mock survey brightest filaments contain a smaller fraction of low-luminosity galaxies.

the properties of these systems might evolve with cosmic time, and wider-area surveys will permit a more detailed appraisal of the very largest detectable systems.

2.8 A closer look at high-luminosity structures

Figure 2.17 best highlights the differences between mock and observed connected structures. We take a more in-depth view of the brightest systems identified in the 2dFGRS catalogue and investigate how these systems are assembled both in mock surveys and the 2dFGRS. For each survey, we determine the distribution of galaxy luminosities that assembled the largest filament. In Figure 2.18 we plot the contribution of each luminosity bin to the total filament luminosity (i.e. the integral of each function should be one), and find the mock (*grey*) surveys do not include the same fraction of "fainter" ($10^{10}\,h^{-2}\,\mathrm{L}_{\odot} \sim L^{\star}$) galaxies found in the 2dFGRS (*black*) superstructure. Despite noise in the mock data arising from fewer galaxies, the 2dFGRS distribution appears to match well to the model at $L > 10^{11}\,h^{-2}\,\mathrm{L}_{\odot}$. These data suggest it is in this luminosity domain

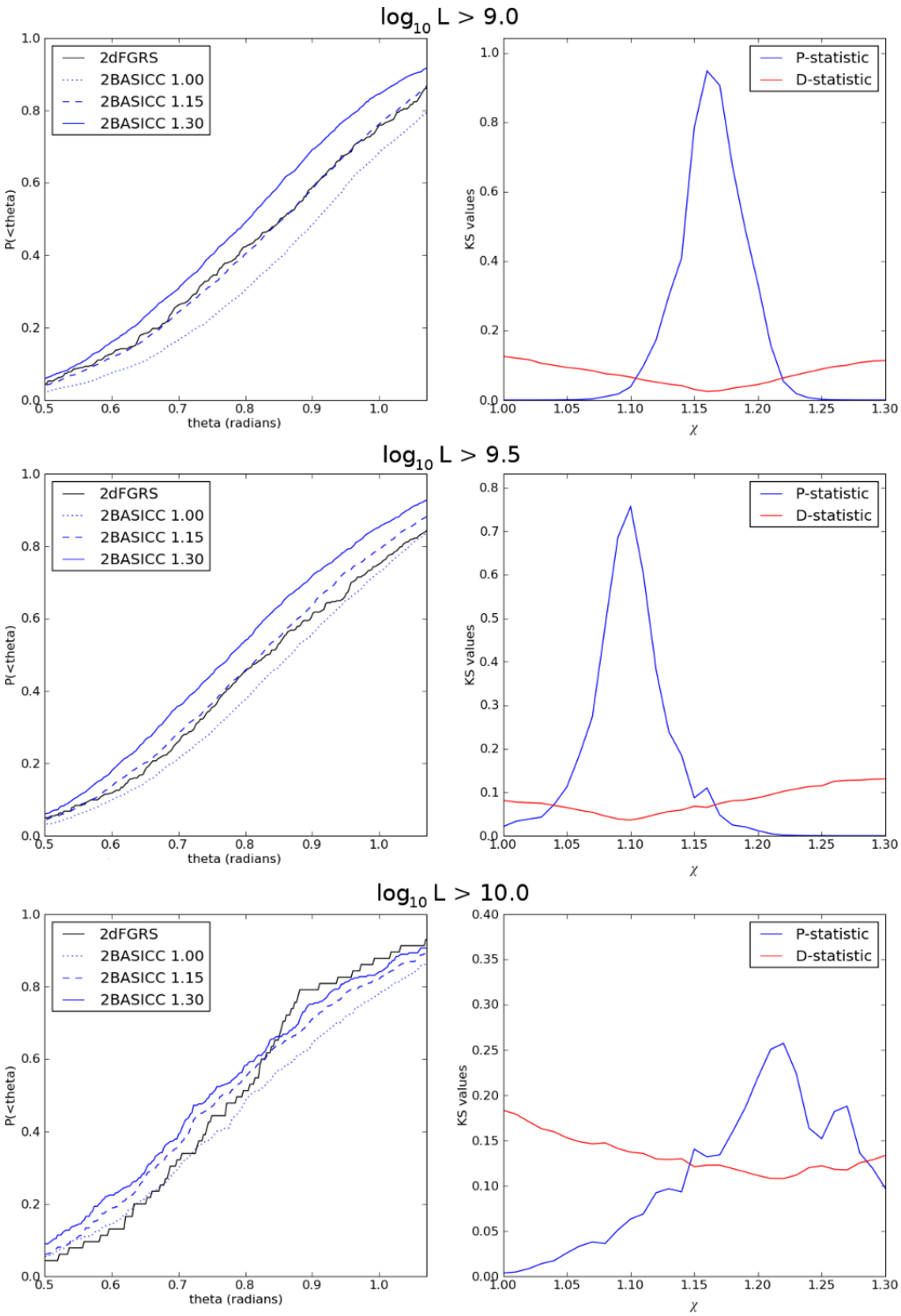

Figure 2.19: Connected system orientation distribution for three luminosity cuts. Trial mock data, using different χ values are shown in the *left* panel. The Kolmogorov-Smirnov statistics used to select the optimal χ value are shown in the *right* panel, where good fits to the 2dFGRS data are indicated by low D values (red) and high P values (blue).

that mock and observed filaments are most similar. To investigate both this and low-luminosity differences, we compare the distribution of connected structure luminosities following cuts to the input galaxy luminosities. To ensure a system detected with different luminosity limits still has the same luminosity, we re-correct the galaxy luminosities as described in §2.4.1. We illustrate this correction with an example 2dFGRS galaxy: situated at a redshift $\mathrm{z} = 0.0230$ with weight 1.1, magnitude $\mathrm{b_J} = 14.4$ in a region with limiting magnitude $\mathrm{b_J^{lim}} = 19.415$. At this redshift, the minimum observable luminosity is $10^{8.07}\,h^{-2}\,\mathrm{L}_\odot$, the galaxy has unweighted luminosity $\mathrm{L_{gal}} = 10^{10.07}\,h^{-2}\,\mathrm{L}_\odot$; sources exceeding $\mathrm{L_{max}} = 10^{10.24}\,h^{-2}\,\mathrm{L}_\odot$ are unobserved. We correct for this unobserved luminosity by scaling the luminosity, in this instance by a factor of 1.18. Table 2.2 shows how the correction changes for this particular galaxy as one makes progressively brighter luminosity cuts. With larger cuts the correction increases - our motivation for setting a z=0.12 limit in §2.4.1 arose from keeping this correction factor below two for the most distant galaxies. Because the corrections increase sharply beyond $L^\star$, we cannot make cuts at the $L > 10^{11}\,h^{-2}\,\mathrm{L}_\odot$ level Figure 2.18 suggests is most appropriate.

We apply three luminosity cuts to the input data: $L > 10^9, 10^{9.5}$ and $10^{10}\,h^{-2}\,\mathrm{L}_\odot$. Luminosity corrections assigned to the remaining galaxies, if correctly calculated, should approximately recover the same total survey galaxy luminosity. We calculate the total galaxy luminosity for the two survey regions separately and find, for the above limits, these were (NGP,SGP)=(2.12,3.39),(2.12,3.39) and (2.16,3.46)$\times 10^{16}\,h^{-2}\,\mathrm{L}_\odot$ respectively.

The structure detection algorithm was applied to these survey subsets; for each luminosity cut in the mock analysis a new χ was derived based on the best angular distribution match between the 2dFGRS and the mock structures. Figure 2.19 shows three χ values (originally used in Figure 2.13) for each of the three luminosity cuts (*left* panel). In the *right* panel we show the Kolmogorov-Smirnov test data used to determine the best fit to the 2dFGRS, which corresponds to a low D value coupled with a high probability P. From these data, we adopt best fit values[8]: χ_9,$\chi_{9.5}$,χ_{10}=(1.16,1.10,1.21).

In Figure 2.20, we show the RA-z plots of the detected 2dFGRS connected systems, colour coded by system luminosity (*top* panel) for the three luminosity cuts. In the *bottom* panel, the luminosity distribution functions of the 2dFGRS (black) and 2BASICC (grey)

[8]We note the fit for χ_{10} is not as statistically significant as the others. This is most likely due to the fewer number of systems increasing the noise in the distribution.

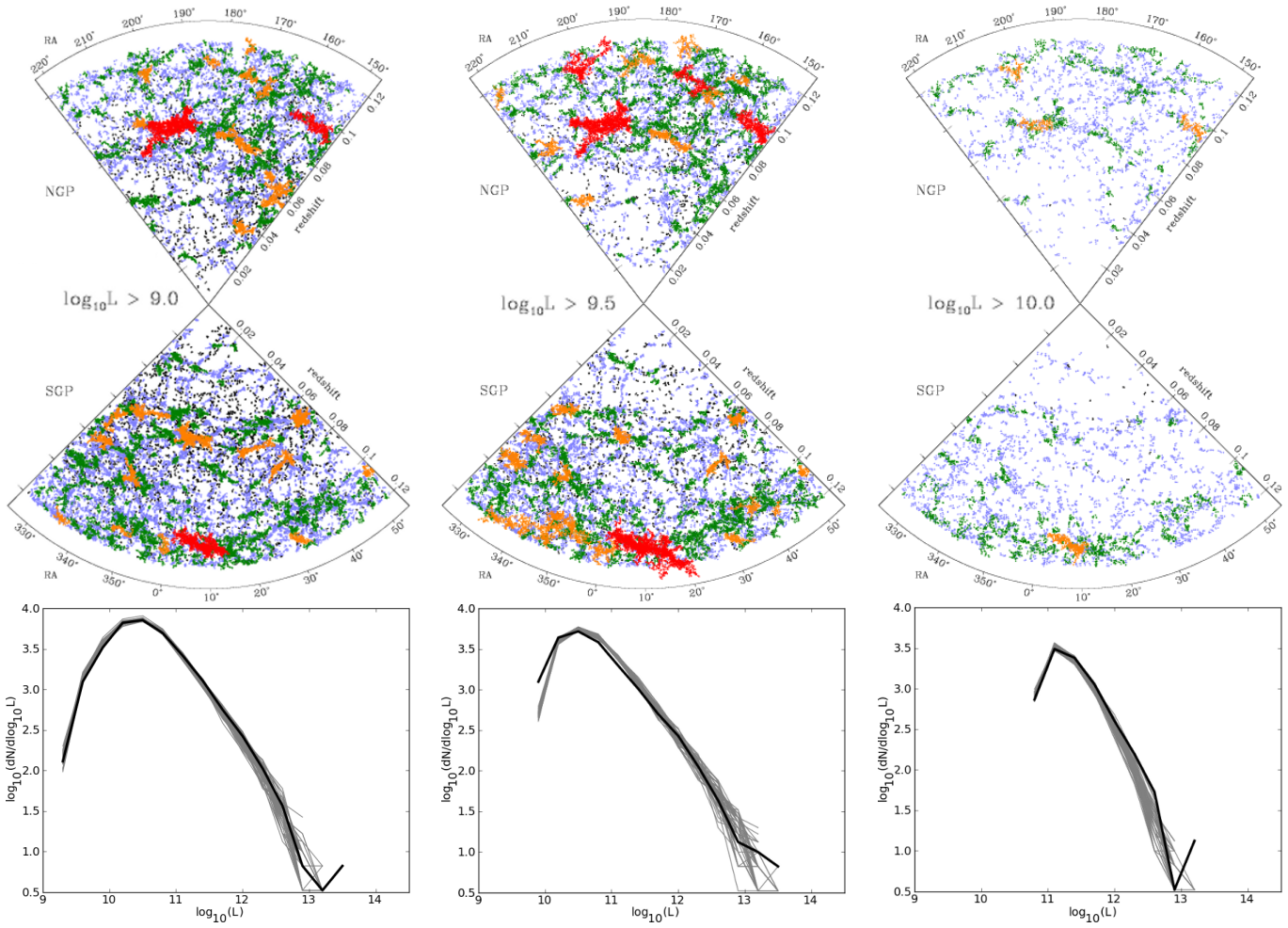

Figure 2.20: (*top*) Connected structures detected in the 2dFGRS for three minimum galaxy luminosities. The colours represent the luminosity of the connected structures. Readers are referred to Figures 2.6 or 2.15 for the legend. (*bottom*) The luminosity distribution of the 2dFGRS connected structures (black line) compared to those of the mock surveys (grey).

$\log_{10}L_{cut}$	Correction factor	Notes
0	1.14	Bright correction only
8.07	1.18	Default for this galaxy
8.5	1.24	
9.0	1.42	
9.5	2.04	
10.0	6.88	
10.1	13.11	
10.15	22.59	Beyond weighted galaxy luminosity
10.22	281.26	
10.24	∞	Faint luminosity limit $>$ bright limit

Table 2.2: The luminosity correction required to account for unobserved flux based on an example 2dFGRS galaxy at z=0.02.

mocks are compared. Whilst there is now better agreement at the faint-end of the distributions, mock surveys are still not able to produce the high-luminosity filaments seen in the 2dFGRS. Galaxies below the luminosity limit in the mock survey are more likely to be taken from high-density environments than those in the 2dFGRS. Whilst this will on average generate lower-luminosity clusters and groups, there should remain more interpolation points with which to connect clusters together. Combined with the lower ejection fractions (through $\chi > 1$), this suggests mock surveys should be better placed to produce bright filaments comparable with the 2dFGRS. That they do not certainly warrants further study.

2.9 Recovering connected structures with photometric redshifts

Motivated by the availability of large-area surveys featuring only photometric redshifts, in this section we evaluate how the structure-finding algorithm performs when applied to data with significant redshift uncertainty. Taking cues from our analysis of photometric redshift data in §3.2.2, we seek to establish what structures could be identified if the 2dFGRS survey relied on photometric as opposed to spectroscopic redshifts.

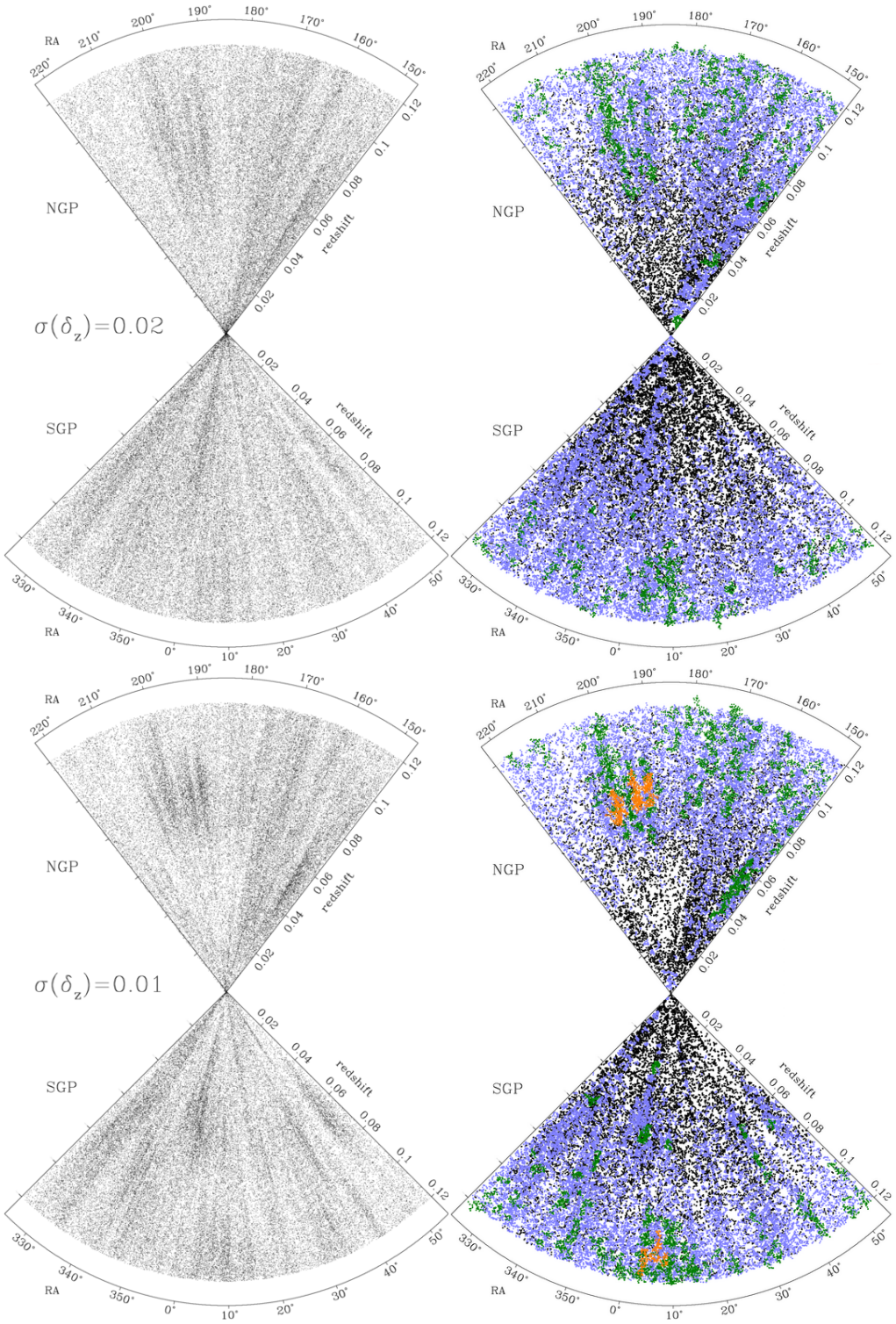

Figure 2.21: The input galaxy catalogues (*left*) and structures detected with the algorithm (*right*) when 2dFGRS redshifts are perturbed randomly by a Gaussian of width σ_{δ_z}=0.02 (*top*) and σ_{δ_z}=0.01 (*bottom*). Colour coding in the *right*-hand plots corresponds to the luminosity of the recovered structures, using the same scale shown in Figures 2.6 and 2.15.

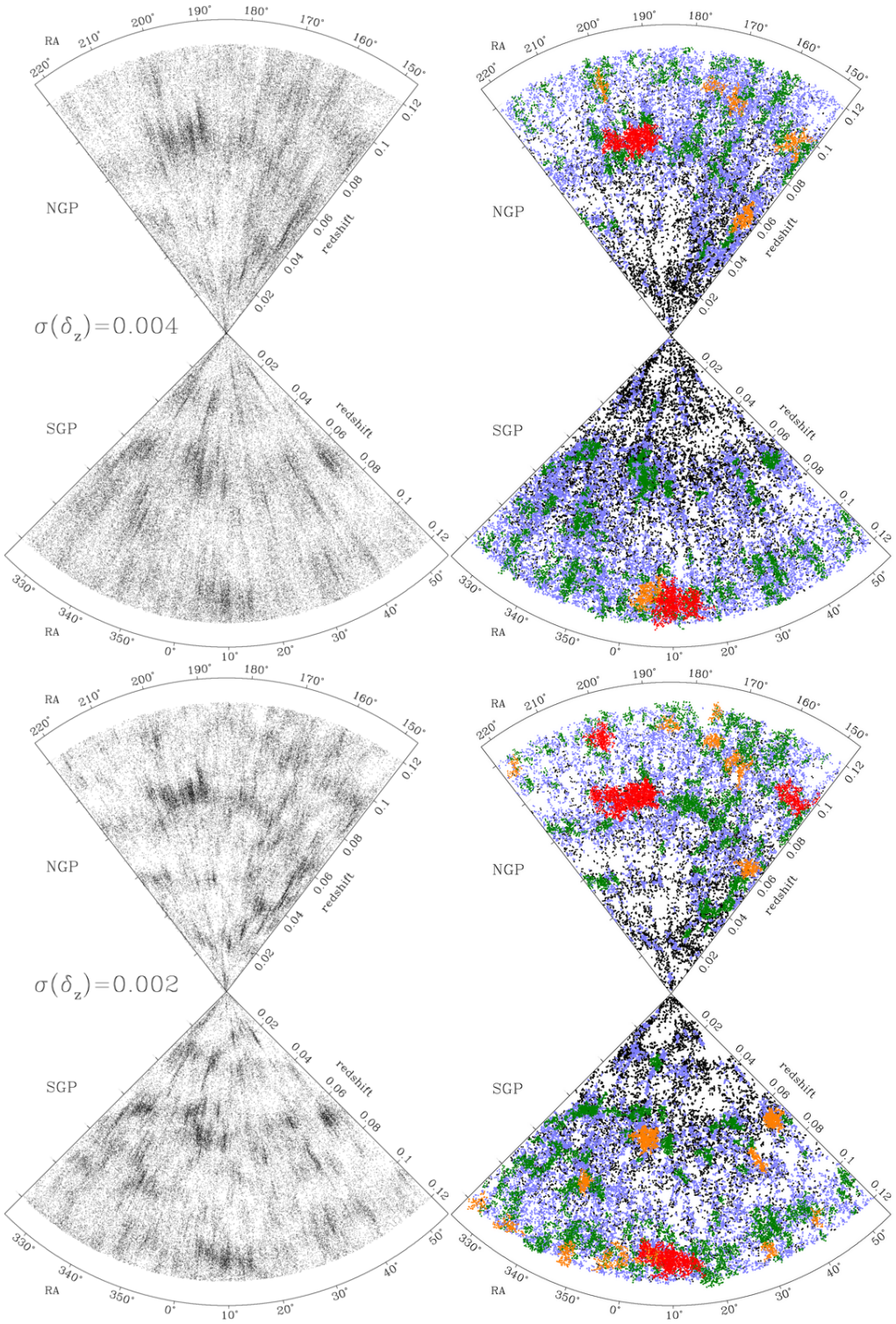

Figure 2.22: The input galaxy catalogue (*left*) and structures detected with the algorithm (*right*) when 2dFGRS redshifts are perturbed randomly by a Gaussian of width σ_{δ_z}=0.004 (*top*) and σ_{δ_z}=0.002 (*bottom*). Colour coding in the *right*-hand plots corresponds to the luminosity of the recovered structures, using the same scale shown in Figures 2.6 and 2.15.

We simulated this scenario by applying a randomisation to the galaxy redshifts in the source 2dFGRS catalogue. Four modified catalogues were created by applying a random perturbation to each galaxy redshift, based on the redshift uncertainty $\delta_{\rm z}({\rm z_s}) = \Delta {\rm z}/1 + {\rm z_{2dFGRS}}$. Under the four scenarios explored here, $\delta_{\rm z}$ was randomly drawn from a zero-mean Gaussian distribution with standard deviations $\sigma_{\delta_{\rm z}}$=0.02, 0.01, 0.004 and 0.002. The latter two are respectively four and eight times smaller than the characteristic redshift uncertainty achieved by the DR7 SDSS photometric redshift algorithm, and so represent what may be possible with improved photometric redshift techniques.

We apply the structure-finding algorithm to each of the four surveys, keeping all detection parameters the same. Figures 2.21-2.22 show the source catalogues in addition to the distribution and luminosities of the detected structures.

In each of these cases, we calculate the connected structure luminosity distribution and compare this to the original 2dFGRS, where the redshift uncertainty is negligible relative to the errors introduced here ($\sim$85km/s). Compared to the filament catalogue in Figure 2.17, the luminosity distributions shown in Figure 2.23 show a more homogenised filament population; erasing the large-scale, large luminosity structures tends to increase the number of median-luminosity systems. Whilst visually the $\delta_{\rm z} = 0.002$ case does appear to identify the large structures detected in the original survey, their luminosities remain significantly attenuated.

These data suggest that, although one may identify the very largest filamentary structures with photometric redshifts, current photoz estimators fail to recover the range of connected structure luminosities observed in the spectroscopic data. Using the spectroscopically selected systems as a benchmark to achieve the same range, photometric redshift uncertainties of less than $\delta_{\rm z} = 0.002$ are required - currently ten times smaller than the capabilities of current photoz estimation codes[9].

We note the results produced here are dependent on the parameters of the algorithm. Although parameters may potentially change to improve the recovery from noisy redshift data, one should compare the systems recovered to those from the equivalent spectroscopic effort. A fairer appraisal of the effect photoz estimates have on these structures could be achieved with closer study of a mock survey. Where the galaxy halo-memberships are known, one may establish the quality of the recovering the underlying

[9]See §3.2.2 for a discussion of photoz algorithm performances

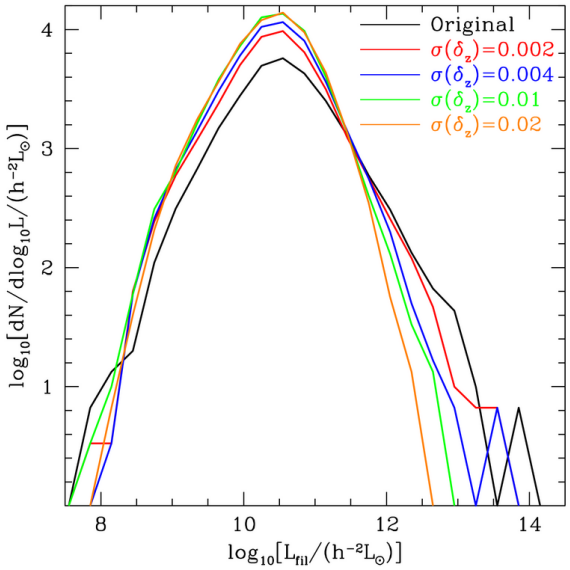

Figure 2.23: The recovered filament luminosity distribution for the four photometric uncertainty scenarios (red, blue and green and orange lines), compared to the original 2dFGRS distribution (black).

mass distribution independent of the performance of the algorithm itself.

2.10 Summary

In this chapter, we have described a simple algorithm with which to define connected structure within galaxy redshift surveys and applied it to the 2dFGRS. This algorithm explicitly addresses the redshift-space distortion associated with rapidly moving galaxies within groups and clusters. The $7,603$ 2dFGRS connected structures at $\mathrm{z} \leq 0.12$ containing at least two members range up to $\sim 200 h^{-1}\mathrm{Mpc}$ in extent, but are mostly associations of two $L^{\star}$ galaxies. Quantifying object sizes via their total luminosities, corrected for the survey flux limits, we find that the largest systems are filamentary in nature and have $\mathrm{b_J}$ luminosities of almost $10^{14}\, h^{-2}\, \mathrm{L}_{\odot}$.

Applying the same algorithm to ΛCDM mock 2dFGRS catalogues, constructed using large-volume dark matter simulations and the semi-analytical model of Baugh et al. (2005), we find a broadly similar distribution of structures to those in the real data. There are, however, a few differences in detail. Many of these result from the fact that the model places too many $L \lesssim L^{\star}$ galaxies into groups and clusters compared with the

2dFGRS. This biases the orientation distribution of the systems containing at least 20 galaxies to contain more radially aligned objects in the mock survey than in the 2dFGRS. Applying a crude correction to the algorithm to enable it to recover the same orientation distribution in the mock survey as it does in the 2dFGRS leads to the largest mock structures being significantly less luminous than those in the 2dFGRS.

It is clear that at least some of the differences between the properties of the structures in the 2dFGRS and the mock catalogues arise from inadequacies in the galaxy formation model that was used to construct the mocks. We have attempted to overcome these inadequacies as far as possible through empirical corrections. Our analysis indicates that the largest filamentary structures seen in the 2dFGRS are not reproduced in the mock catalogues. However, while this discrepancy could signal a failure of the standard ΛCDM cosmological model on large scales, it seems more plausible that it reflects a shortcoming in the predictions of our models of galaxy formation for the abundance and spatial distribution of galaxies on small scales.

In comparing the connected structure luminosity distribution between mocks with different cosmologies and galaxy formation models, we find they are broadly similar. This arises from our choice of different χ values required to emulate the 2dFGRS structure orientation distribution. This suggests the structures detected, through the χ parameter required to match observed data, may potentially be used as probes of cosmology or galaxy formation. With larger surveys, one may better investigate the high-mass tail of this distribution, which may be more sensitive to cosmology.

Our findings in the final section suggest that at least an order-of-magnitude improvement is needed in current photometric redshift techniques before they can recover the range of structures identified with spectroscopic data. Although one may still recover the very largest structures in the survey with photometric redshifts, to extract less luminous systems an alternative approach to exploit deep photometric imaging is required.

Chapter 3

ORCA: The Overdense Red-sequence Cluster Algorithm

Preface: Galaxy Clusters

One way of identifying large scale structure is by the detection of groups and clusters of galaxies. Galaxy clusters are integral tools in our drive to test the ΛCDM cosmological model and our understanding of galaxy formation. Their redshift evolution imposes constraints on the dark matter content of the universe, cosmic mass density and the growth of density fluctuations. The deep potential wells of clusters provide laboratories to study in-situ gas-galaxy interactions and through lensing, access to conditions in the early universe. Assembled for a significant fraction of the star-forming history of the universe, clusters may also provide insights into how environmental effects shape the evolutionary path of galaxies. Moreover, measuring this assembly over cosmic time provides a probe of dark energy complementary to Cosmic Microwave Background (CMB) studies. From such cluster science, we are able to piece together both the origins of the universe, and how it evolved over time.

Clusters lie at the intersections of filaments, and trace out the high-density peaks in the matter distribution of the universe. They have a variety of important astronomical applications, and may be used as points of interpolation with which to delineate the cosmic web. In the previous chapter, we used groups and clusters of galaxies as the "building blocks" of the connected structure we seek to characterise. We turn, in this chapter to clusters themselves as the focus of our study.

3.1 Overview

This chapter presents a cluster finding algorithm that seeks to remove the reliance similar algorithms have on two aspects of detection: model descriptions of cluster red sequence positions, and parametric distributions of cluster members on the sky. Our approach involves a scan of colour-magnitude space to locate cluster sequences coupled with the Voronoi Tessellation technique to estimate the surface density distribution. Whilst algorithms have in the past used Voronoi Tessellations to find clusters, previous attempts either do not exploit the red sequence or instead use (time-consuming) photometric redshift distribution functions that rely sensitively on accurate photometry, the absolute calibration and number of photometric bands. In this chapter we describe the algorithm and apply it to two datasets. The first is a $\sim$7 square-degree mock survey based on a Pan-STARRS lightcone. The second dataset is a mock Medium Deep Survey (MDS) Pan-STARRS catalogue, based on the ΛCDM model and a semi-analytic galaxy formation recipe. Knowledge of galaxy-halo memberships in the mock allows a quantification of algorithm performance. Because our development of the cluster-finder was based primarily around the observational data, we present the algorithm first with our findings from that survey, followed by our analysis of the mock data.

The outline of this chapter is as follows. In section §3.2 we present the data we use in this study. Section §3.3 describes the algorithm step-by-step. Section §3.4 describes the application and testing of the algorithm using real astronomical data, followed by a brief comparison to existing data in §3.5. §3.6 describes the detection of mock clusters followed by performance tests on the simulated catalogues in §3.7. In section §3.8 we summarise our findings.

3.2 The source galaxy catalogues

In this section we describe briefly the two datasets analysed in this chapter, in addition to any further selection criteria made.

3.2.1 SDSS Stripe 82 data

We extract Sloan Digital Sky Survey Data Release 7 *griz* photometry for all sources with extinction-corrected (Schlegel et al., 1998) *r*-band model magnitudes $r \leq 24$ in the deep coadd stripe centred on the celestial equator ("Stripe 82") from the SDSS Catalog Archive Server (CAS[1]). To minimise stellar contamination, we select only galaxies where the offset between the *r*-band PSF and model magnitudes satisfies $|r_{\rm PSF} - r_{\rm model}| > 0.05$. We exclude bright ($r_{model} < 14$) galaxies and spurious sources such as overly de-blended galaxies and fragmented stellar haloes. The latter selections were achieved by requiring the following set of constraints in the CAS SQL query to hold:

- `BINNED1` or `BINNED2` or `BINNED4` > 0
- `BLENDED` or `NODEBLEND` or `CHILD` != `BLENDED`
- `EDGE` or `SATURATED` $= 0$

Full details of the query and the keyword definitions can be found in Appendix B. The resulting catalogue generated from this query had fewer stellar contaminants than merely selecting objects from the SDSS `Galaxy` table. Figure 3.1 shows some examples of the systems omitted by the query above that would otherwise have been included. From visual impressions, a large fraction of these arise from the mis-identification of spurious noise around stellar haloes.

Although no spectroscopic or photometric redshift estimates are used in detections, we post-process the cluster catalogue to estimate the redshift of each system. Cluster galaxies are assigned spectroscopic redshifts by matching source positions in the SDSS DR7, WiggleZ DR1 (Drinkwater et al., 2010) and 2SLAQ (Croom et al., 2009) catalogues to within $1''$. Where spectroscopic redshift data is unavailable, we use SDSS DR7 photometric redshifts (see Abazajian et al., 2009, and references therein). To increase both the source catalogue redshift completeness and the redshift accuracy for galaxies with no spectra, we supplement these data with additional photometric redshifts. We select all galaxies later identified by `ORCA` in the Stripe 82 catalogue described in chapter 4 and estimate their redshifts using the `hyperz` code[2] (Bolzonella et al., 2000) with *ugriz*

[1] `http://casjobs.sdss.org`
[2] `http://webast.ast.obs-mip.fr/hyperz`

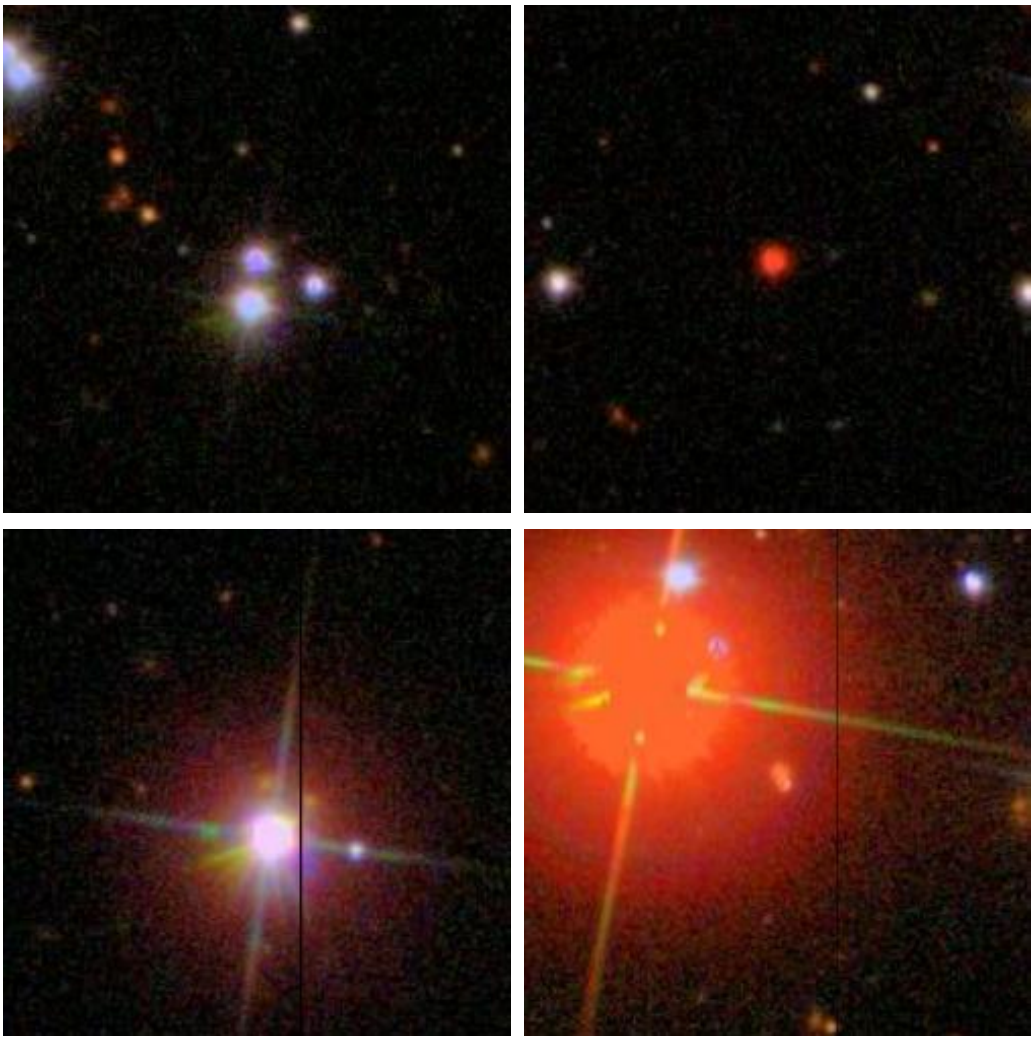

Figure 3.1: Examples of sources from the SDSS catalogue that were identified in the "Galaxy" table but subsequently rejected by our SQL query. Each candidate is at the centre of the image. The *top* row are stellar contaminants, whilst the *bottom* row are spurious stellar haloes misidentified as genuine astronomical sources.

model magnitudes and errors. Full details of this photometric redshift catalogue can be found in the next section. The SDSS Stripe 82 input catalogue contains 11,358,087 galaxies with Galactic extinction corrected (Schlegel et al., 1998) *griz* model magnitudes, over $-50^\circ < \alpha < 59^\circ$ and $\delta = \pm 1.25^\circ$. In this study, we concentrate on a 7 square degree subregion within this catalogue, centred at $(\alpha, \delta) = (355.52^\circ, 0^\circ)$ comprising 291,389 galaxies (magnitude cuts applied to these galaxies for cluster detection are discussed in §3.3.7).

3.2.2 Calculating and evaluating photometric redshifts

The estimation of galaxy redshifts from broad-band colours is set to become central to future efforts in large-scale cosmology. For example BAO and cluster distance estimates will play important roles in constraining the nature of the universe, and the Dark Energy Survey (DES) will investigate both of these with deep five-band photometry. Similarly, with only limited spectroscopic coverage of extremely wide-area datasets such as the Pan-STARRS 3π survey, much of the planned extragalactic science is derived from photometric redshift estimates. It is therefore important that these redshifts are accurate and robust.

To demonstrate the extraction of photometric redshifts from a large photometric sample, we use the freely available `hyperz` (Bolzonella et al., 2000) template fitting code. This package uses seven GISELL (Bruzual and Charlot, 1993) SEDs. Our source data, a sample of $\sim 60,000$ early-type SDSS Stripe 82 cluster members from the `ORCA` cluster detector introduced in this chapter, features deep *ugriz* SDSS model (considered optimal measurements of galaxy flux) AB magnitudes. For each galaxy, the SDSS photometric pipeline fits two models to the 2D galaxy image - a de Vaucouleurs profile (de Vaucouleurs, 1953b) best suited to early types or central bulges in disk galaxies, and an exponential profile better matched to late-type galaxies. Flux is then measured within this profile.

`Hyperz` operates by reading configuration parameters and input data from a variety of files stored on disk. We modified the source code to permit command-line queries accepting input magnitudes and uncertainties for a single galaxy, returning the results to the screen. Not only does this reduce the read/write frequency to disk, it also allows the development of a parallelised pipeline to process the source data. By making these adjustments, our photometric redshift calculations were complete in 18 hours, compared to approximately 5 days with the original (serial) implementation. For our galaxy sample, we search in a redshift range $0.0 \leq z \leq 0.8$. Because galaxies in this sample are cluster members, we use only three of the SED templates: E, S0 and Sa. `Hyperz` includes routines for attenuating the emergent spectrum with dust from within the galaxy, and in this investigation we applied the Calzetti et al. (2000) reddening law, searching for fits between 0 and 1 magnitudes in 11 steps. From our 60,000 cluster galaxies, 1549 galaxies

also have spectroscopic redshifts - it is these we use to infer the performance of the algorithm. The dataset will be used in our analysis of galaxy clusters both in this chapter, and more extensively in chapter 4. For each spectroscopically measured galaxy, there is also a photometric redshift estimate from the SDSS pipeline (as described in Abazajian et al., 2009, §4.6). For both SDSS and the study here, we plot in Figure 3.2 the correlation between photometric (z_p) and spectroscopic (z_s) redshifts, as well as $\delta_z(z_s) = (z_s - z_p)/1+z_s$. Our uncertainty estimates, δ_z, are relative to $1+z_s$ so as to reflect the higher impact errors have at low redshifts.

From Figure 3.2, it is evident the `hyperz` approach does not perform as well as the SDSS algorithm. The error dispersion in the *bottom left* panel suggests that `hyperz` systematically underestimates the actual redshift. Whilst the scatter appears similar for the two studies, the 1-σ width (3-σ clipped) of `hyperz` redshift estimates is σ_{δ_z}=0.029, compared to σ_{δ_z}=0.016 for the SDSS dataset. This latter result compares favourably to photometric redshift estimates of the COMBO-17 survey, where Wolf et al. (2004) report errors between 1-2%, based on *UBVRI* photometry complemented with an additional 14 optical filters. The `hyperz` "hole" at $z_p, z_s \simeq 0.3$, similarly evident in the analysis of Cai et al. (2009), appears to suggest the algorithm converges to the wrong solution for this redshift regime. In our subsequent analysis of clusters using this data, we note that this incompleteness does not significantly influence our findings. Although this effect is not endemic to all `hyperz` analyses, this demonstrates the sensitivity of the process to the input parameters applied - without in-depth knowledge of the galaxy selection function, the level of accuracy will remain low. In spite of this, with the exception of projects such as SDSS3:BOSS and BigBOSS[3], a large number of planned or ongoing projects will provide wide, deep coverage of the sky without spectroscopic support. To maximise large-scale structure science from DES, Pan-STARRS, LSST and *Euclid*, one must produce well-defined, well-characterised datasets capable of producing low-scatter photometric redshifts. One alternative, to de-project the distribution of galaxies without direct redshift calculations, is explored in this chapter.

[3] `http://bigboss.lbl.gov`

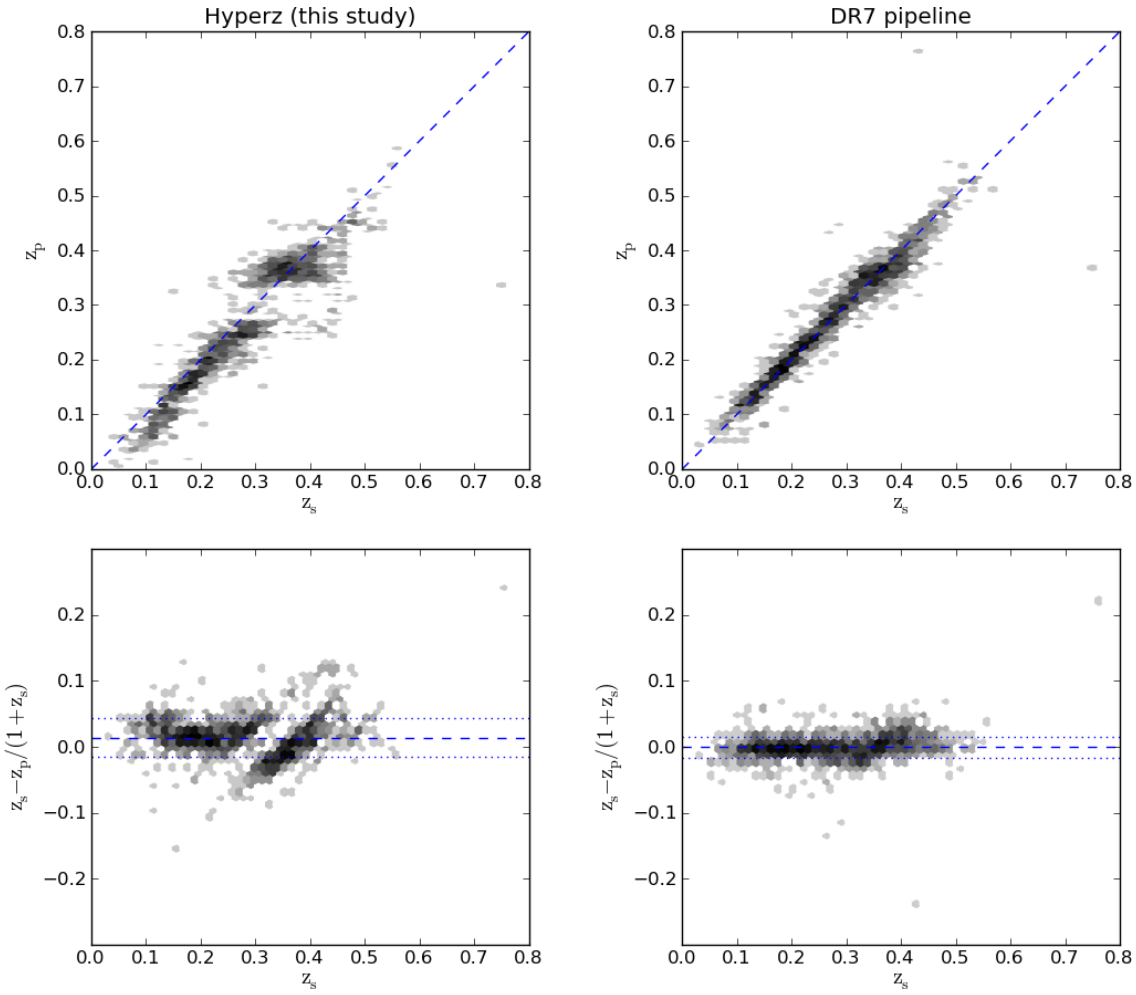

Figure 3.2: (*top*) Spectroscopically determined redshifts compared to their photometric estimates for 1549 galaxies in the ORCA cluster galaxy sample, and the error dispersion for the above data (*bottom*). Dotted blue lines indicate $\pm 1\sigma$ over the whole redshift range, following a 3-σ clip of outliers.

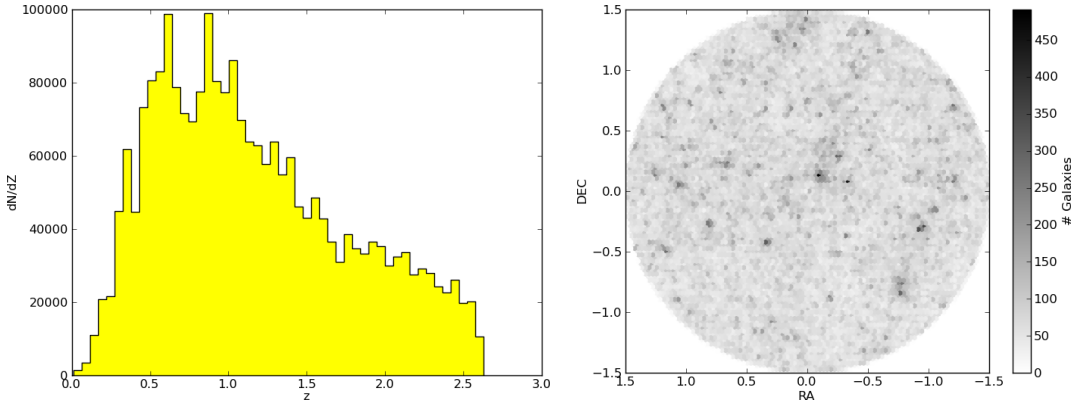

Figure 3.3: (*left*) The redshift distribution of galaxies and groups from the MDS lightcone. (*right*) A two-dimensional histogram showing the projected density distribution of galaxies out to z~ 0.7.

3.2.3 Pan-STARRS MDS mock galaxies

Cai et al. (2009) discuss the assembly of a light cone from the Millennium Simulation (Springel et al., 2005) with a 3° opening angle, equivalent to a single pointing of the Pan-STARRS Telescope 1 (PS-1) Medium Deep Survey (MDS). The Millennium Simulation provides the ΛCDM architecture into which galaxies are populated using the Bower et al. (2006) semi analytic `GALFORM` model (Cole et al., 2000). This creates a dataset with PS-1 *grizy* photometry for 2,346,468 galaxies down to a magnitude limit of $r < 27.5$ (equivalent to the expected 5σ depth for the PS-1 MDS) and a median redshift of $z = 1.05$. Figure 3.3 shows plots of the projected density of mock galaxies and their redshift distribution. Several overdense regions are already apparent in the density plot. One challenge of the algorithm is to separate chance projections from genuine clusters.

3.3 The algorithm

Here we describe the main steps of the `ORCA` algorithm.

3.3.1 Algorithm Outline

Applying the `ORCA` algorithm to an example survey with two colours can be summarised as follows:

1 We define a simple photometric selection using the colours and magnitudes of

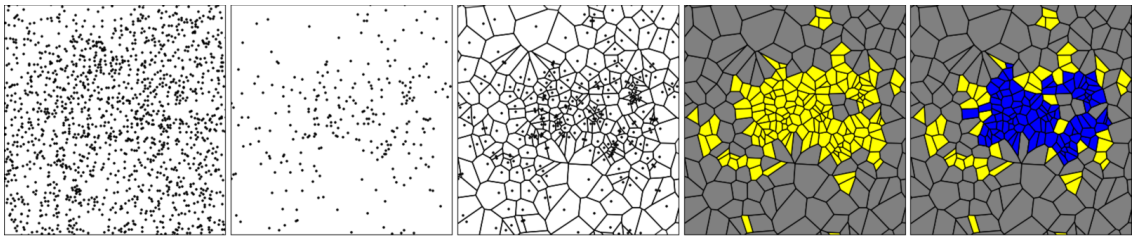

Figure 3.4: A depiction of the ORCA detector applied to a 9'x9' cut-out region of Stripe 82. Starting with all galaxies in the box (*first* panel), a photometric filter (§3.3.2) isolates galaxies within a specific redshift range (*second* panel); any clusters in this field will be evident as surface overdensities. In the *third* panel, we compute the Voronoi diagram (§3.3.4) of the distribution to estimate the surface density of remaining galaxies. These are separated into overdense (yellow) and underdense (grey) cells in *panel* four according to how likely they are to belong to a random distribution (§3.3.4). In the *final* panel, we use a Friends-Of-Friends percolation algorithm (§3.3.5) to connect overdense cells until the density of the whole system falls below a density threshold. Galaxies in the blue cells become members of a cluster if there are at least $N_{\rm min}$ linked members.

the sample. This selection could be simple, for example narrow slice(s) in colour-magnitude space(s), or an arbitrary selection function. This selection function can be modified in successive applications of the algorithm to blindly scan the full photometric space, and thus isolate red-sequences across a range of redshifts.

2 In each pass of the algorithm, we apply the photometric selection to the catalogue, thus greatly restricting the total number of galaxies under consideration. In the case of using two colours concurrently, this can be a very effective means of reducing fore- and background contamination to a putative cluster characterised by some red-sequence.

3 After the selection, we calculate the Voronoi diagram of the projected distribution of galaxies on the sky. The inverse of the area of each convex hull surrounding each galaxy can be used as an estimate of the local surface density.

4 Galaxies residing in dense cells (satisfying some threshold criteria) can be connected together into conglomerations. If enough galaxies are joined together in this way, we define a cluster.

5 In the blind scan, successive photometric cuts may select the same structures (since the adjustment of the selection is by design less than the typical width of a red-sequence). Multiple detections of the same structure are identified and reduced to a single detection (we discuss how this was implemented in §3.3.6).

A graphical overview of the above procedure can be seen in Figure 3.4. Gladders and Yee (2000) and Gladders and Yee (2005) show that using a set of colour filters allows one to scan the expected cluster red sequence over a range of redshifts. This study will demonstrate that combining a sweeping set of colour selections with a Voronoi Tessellation density estimator is a powerful way of identifying clusters of galaxies. Our method does not rely on any specific spectral features other than the 4000Å break ubiquitous in early type galaxies. With the sampling technique we use, clusters can be detected in multiple photometric filters. These multiple detections are merged by picking the 'best' detection; we discuss how this was implemented in §3.3.6.

3.3.2 Photometric filtering

In large-scale imaging surveys, groups and clusters are apparent as overdensities in the projected distribution of galaxies. Cluster detection methods reliant only on determining the projected galaxy density distribution are often plagued by two problems: (i) projection effects contaminating clusters with unassociated foreground and background galaxies (ii) the inclusion of spurious cluster detections arising from noisy data or chance overdensities.

To mitigate these problems, the contrast of genuine clusters can be enhanced by applying a selection filter in colour-magnitude space, to isolate the red-sequence ridge-line. We parametrise our selection filter as a slice in colour-magnitude space, defined by a colour-magnitude normalisation ($c_{\rm m20}$, the colour at twentieth magnitude), slope $\beta(c_{m20})$ and width $\sigma(c_{m20})$. The expected evolution of red sequence colours is constrained from simple stellar evolution models, meaning an appropriate set of filters allows, by scanning over a set of these filters, the isolation of clusters over a slew of redshifts. Figure

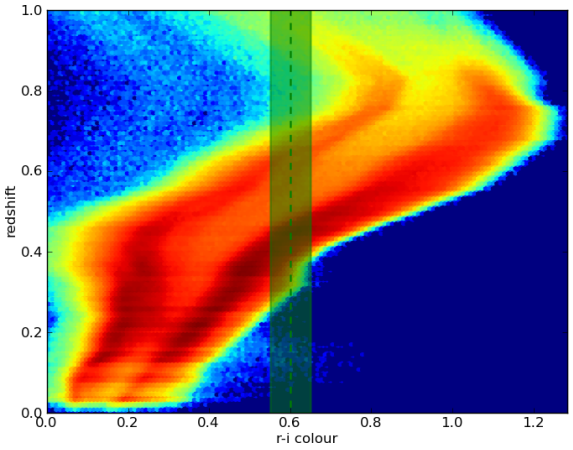

Figure 3.5: The redshift evolution of the observed-frame *r-i* colour from a sample of mock galaxies. The colours indicate the density of galaxies at each point, with red being the highest. We are able to exploit this observed relation to isolate cluster galaxies within a specific redshift range by using a selection filter (such as the shaded strip in this Figure) to select galaxies from a narrow colour range.

3.5 shows the redshift evolution of galaxy colours in a sample of mock galaxies (A. Merson, priv. communication) and shows an additional advantage in using such filters. The two tracks visibly demonstrate the bimodality in galaxy colour that manifests itself as the "red sequence" (lower track; Bower et al., 1992) and "blue cloud" (upper track). By selecting galaxies within specific colour range Δc (as denoted by the green region in the Figure), one may isolate red sequence cluster galaxies within the redshift range Δz. Contaminants in this selection are bluer galaxies from higher redshifts. By simultaneously selecting in two filters, one can eliminate degeneracies between colour tracks. We discuss this further in the following section.

The algorithm allows β_{cm20} and σ_{cm20} to adopt any values as the detector scans through colour-magnitude space. By far the simplest prescription, and the one we adopt here, is for a fixed slope and width with normalisation. Although we note evidence for a redshift evolution in the observed-frame sequence slope (Stott et al., 2009; Hao et al., 2009), an appropriate choice of filter width (discussed in §3.3.7) selects sequences with a range of gradients. By taking measurements from a large ORCA cluster catalogue, future refinements to the algorithm may include a varying slope. The values adopted for β and

σ are discussed in §3.3.7.

We scan through colour-magnitude space in a colour $C_{\rm A}$ from blue to red, placing down a series of M filters $f(C_{A_1}), f(C_{A_2})...f(C_{A_M})$ by increasing the normalisation $c_{\rm m20}$ in small increments dc. The size of this increment, set in §3.3.7, allows adjacent filters to overlap, ensuring clusters close to the boundary of a filter are well sampled. Because each filter isolates cluster galaxies (where they exist) from a specific redshift range, the detector can identify multiple clusters in the same line of sight. We determine the sensitivity of the algorithm to projection in §3.4.6.

3.3.3 Dual-colour photometric filtering

Although only one colour is necessary to detect clusters, Figure 3.5 notes the colour-redshift degeneracy apparent in attempting to isolate a redshift regime from a single colour selection. One can break the degeneracy and further reduce the field galaxy contamination by identifying the colour range cluster members have in a second colour $C_{\rm B}$, and subsequently applying a series of joint filters in both $C_{\rm A}$ and $C_{\rm B}$. To establish the $C_{\rm B}$ colour range to scan, we take all cluster members from the preliminary detection ($C_{\rm A}$ only), de-trend their sequence slopes and fit a Gaussian to the colour distribution. The $C_{\rm B}$ colour range $\Delta C_{\rm B}$ is taken to be $\pm 1\sigma$ from the Gaussian mean.

If the Gaussian fit is poor, detection of a clear sequence in both $C_{\rm B}$ and $C_{\rm A}$ is less likely. In this case $\Delta C_{\rm B}$ is simply $\pm 1\sigma$ from the median of the $C_{\rm B}$ colour distribution. The algorithm then scans over this second colour range and attempts to detect the cluster in both colours.

A filter pair in $C_{\rm A}$ and $C_{\rm B}$ (hereafter $\{C_{\rm A}$,$C_{\rm B}\}$) requires a detectable sequence in both colours, and amplifies the cluster signal by eliminating field galaxies in the $C_{\rm A}$ filter that fail to appear within the $C_{\rm B}$ filter. Any cluster in the final catalogue detected in $C_{\rm A}$ must therefore also have been detected in $C_{\rm B}$. This improves the robustness of the algorithm and the reduction of contaminants from spurious detections. Because sub-filters overlap in $C_{\rm B}$ colour-magnitude space, the same cluster may be detected in multiple filters. We apply the prescription described in §3.3.6 to identify and merge clusters that have been detected in more than one filter. The number of filters used to sample any colour range depends on the sampling interval dc set in §3.3.7.

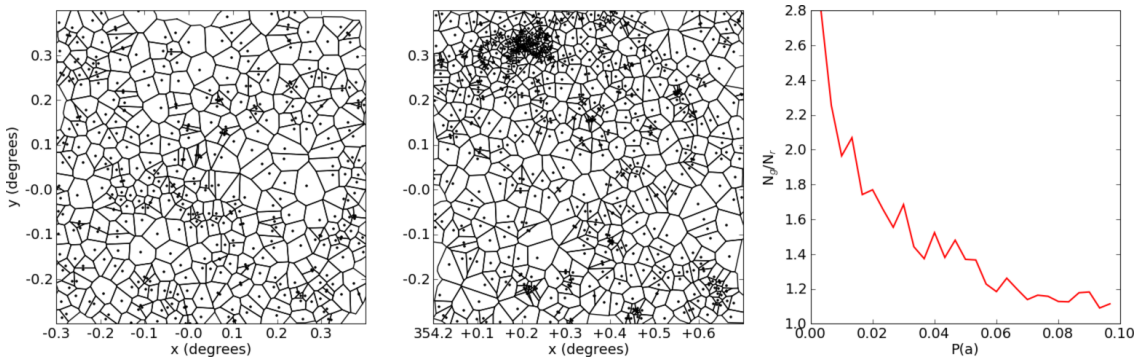

Figure 3.6: An illustration of the Voronoi technique described in §3.3.4. The *(left)* panel is the Voronoi diagram of a random distribution of points. The *(middle)* panel is the equivalent diagram for galaxies in a field with the same mean density as the random field. The *(right)* panel shows the ratio of galaxy cell counts to random cell counts for a range of values of the integral distribution of cell areas (Equation 3.1, derived from Equation 5 in Kiang, 1966). There is a notable excess fraction of galaxy cells relative to random cells at low values of P(a), permitting the use of a threshold to separate clustered galaxies from field galaxies.

3.3.4 Identifying overdensities with the Voronoi tessellation

After increasing a cluster's detectability by suppressing field galaxies with photometric filters, the next step is to calculate the local surface density of each galaxy. Galaxies residing in common regions of enhanced density can then be grouped together into clusters. To quantify the surface density field, we divide the galaxies into Voronoi cells using `qhull`[4] (Barber et al., 1996). The Voronoi diagram is a tessellation of convex hulls, or cells, with each galaxy occupying only one cell. All positions inside a given cell are closer to the cell's nucleus (the galaxy) than any other. Unlike many other detection techniques, the Voronoi tessellation (for VT cluster detection, see Ebeling and Wiedenmann, 1993; Ramella et al., 2001) does not smooth the data, is robust to cluster ellipticity (Plionis et al., 1991), the inclusion of spurious sources, and to local incompleteness arising from camera defects and excised bright stars. The *left* and *middle* panels of Figure 3.6 show Voronoi diagrams with identical mean densities $\bar{\Sigma}$, but with a random point

[4] `http://www.qhull.org`

distribution and a population of galaxies. Galaxies in more concentrated regions tend to have smaller cells.

We define the reciprocal of the galaxy cell area (a_g) as an estimate of the galaxy's local surface density $\hat{\Sigma}_g$. Searching for connected regions of high density identifies statistically significant structures. To determine if a galaxy resides in a high density region of the survey, we evaluate the statistical significance of finding a cell of area a_g in a random field with mean cell area $\bar{a}_{\mathbb{R}}$. We use the Kiang (1966) cumulative function for a Poissonian distribution of points:

$$P(a) = \int_0^a \mathrm{d}p = 1 - e^{-4a}\left(\frac{32a^3}{3} + 8a^2 + 4a + 1\right) \tag{3.1}$$

where $a = (a_g/\bar{a}_{\mathbb{R}})$. We investigate this equation, and possible improvements to the model it describes in chapter 6. The *right* panel of Figure 3.6 shows the distribution P(a) for cells in an example galaxy field relative to a Poisson distribution of the same field size and number of points. Candidate cluster galaxies residing in overdense regions can be selected by cell areas statistically unlikely to arise in a random distribution. An excess of galaxy cells is apparent for low P(a) compared to the random distribution. We identify all galaxies with $P(a_g) < \mathrm{P}_{\mathrm{thresh}}$ in order to select a population of clustered galaxies. The choice of overdensity probability threshold is discussed in §9.

3.3.5 Connecting overdense regions to form clusters

Remaining galaxies belonging only to overdense cells are now grouped together to form clusters. We achieve this by applying a Friends-Of-Friends algorithm to these cells. Rather than a distance criterion, we define a "friend" as an adjacent Voronoi cell sharing at least one vertex. Potential clusters are seeded by ordering the cells with decreasing density, iterating through and connecting adjacent cells. These overdense regions grow by percolation until either no more adjacent overdense cells remain, or the mean cell density of the putative cluster:

$$\bar{\Sigma}_{cells} = N_{gal}\sum_{i=1}^{N_{gal}}\frac{1}{a_i} < \Sigma_{crit} \tag{3.2}$$

Groups of connected galaxies are classified as clusters if they have $N_{gal} \geq N_{min}$. The choice of the critical density threshold Σ_{crit} and N_{min} algorithm parameters is discussed in §3.3.7.

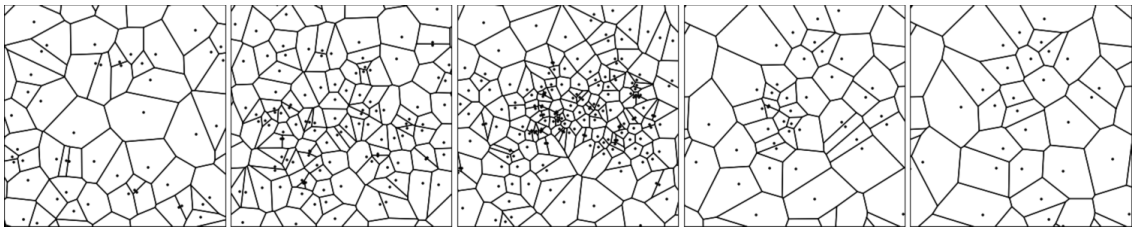

Figure 3.7: A sequence of Voronoi diagrams generated from galaxies in the same area of sky, but selected from different photometric filters. A cluster signal is apparent for some filters, but is not apparent in others. This demonstrates the power of colour selection in isolating galaxies at specific redshifts. In cases where a cluster may be detected in more than one filter (such as the borderline detection in the second panel), the algorithm must decide which cluster to select. This aspect of the detector is discussed in §3.3.6.

3.3.6 Producing a cluster catalogue

In §3.3.2 and §3.3.3 we noted that adjacent photometric filters applied to the input catalogue overlap in colour-magnitude space. With this sampling strategy, the same cluster could be detected in multiple filters. Figure 3.7 shows a sequence of Voronoi tessellations applied to the same area of sky under photometric filters sensitive to different redshift ranges. Because colour scans sample the colour range of a red-sequence at a specific redshift, the cluster will be detected in multiple scans (with a peak contrast where the selection is most effective). In cases of clusters detected multiple times in different photometric filters, the "best" cluster is identified and added to the final cluster catalogue.

For two candidates to be considered detections of the same system, they must have sufficiently similar spatial positions, red sequence fits and cluster members. We quantify the similarity in cluster sequences using linear fits to the colour-magnitude relation for the galaxies in each cluster detected. Sequence slopes can in principle adopt any value permitted by the width of the photometric filter (defined here as σ_f) it was selected in. We quantify the similarity between two sequences with the following criteria:

- *Sequence match 1* (ΔS_1): True if the sequence separation is $< 0.5\sigma_f$ in colour for at least 25% of the magnitude range $m_{\rm BCG} \leq m \leq m_{\rm BCG} + 5$.

- *Sequence match 2* (ΔS_2): True if the sequence separation is $< \sigma_f$ in colour difference for at least 50% of the range described in ΔS_1.

- *Sequence match 3* (ΔS_3): True if the colour difference at $20^{\rm th}$ magnitude, ($\Delta c_{\rm m20}$) between the two sequences is $< \sigma_f$.

- *Sequence match 4* (ΔS_4): True if the clusters were detected in adjacent (overlapping) filters.

To define the similarity in cluster membership, spatial position and extent, we describe the *common-galaxy* fraction and *projection extent* for two clusters, CL_1 and CL_2:

- *Common galaxies* ($cg_{1,2}$): the fraction of galaxies in CL_1 that also belong to CL_2. Similarly, $cg_{2,1}$ is the fraction of CL_2 galaxies also appearing in CL_1. The BCG$_{id}$ boolean notes when clusters share the same BCG.

- *Projection extent* ($pe_{1,2}$): the fraction of galaxies in CL_1 that lie within the Voronoi cell boundaries of the CL_2 cluster. As with cg, $pe_{2,1}$ is the case for CL_2.

With these measures, five tests of "cluster similarity" were devised (Table 3.1). A pair of clusters must pass *at least one* to be considered detections of the same system. Each of these tests account both for the spatial and colour characteristics of the clusters. Because no merging can proceed purely by colour similarity or spatial coincidence, this ensures the separation of associated but distinct systems, and clusters in projection. We balance these requirements with the need to prevent multiple instances of the same cluster appearing in the final catalogue. Where matches between two clusters exist, the thresholds in Table 3.1 make it likely the two systems will be merged. Figure 3.8 shows an example of this process for clusters detected over a number of photometric filters. The multiple detections, conspicuous in the *left* panel as concentric rings, are processed to form the merged catalogue in the *right* panel.

To define the "best" cluster from a list of candidates, we pick out the system with the largest *reduced flux* - the total flux (in the detected band) of all but the three brightest cluster members. This prevents the selection of a cluster including one or two bright galaxies that may not be genuine members, but also makes the choice of best cluster largely independent of the BCG. Once the "best" cluster is selected, the remaining candidates are

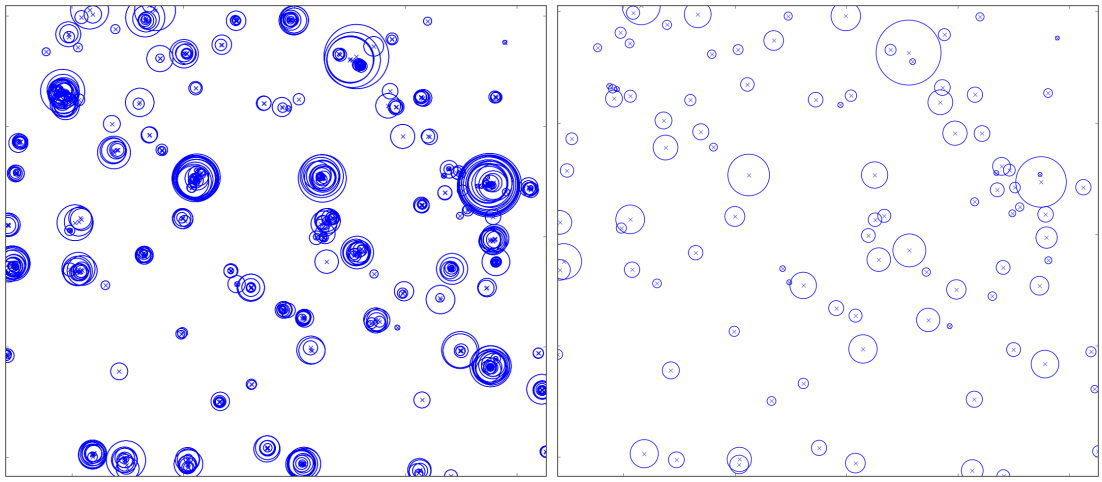

Figure 3.8: An example of the merging of multiple detections in a cluster field. The blue crosses indicate the estimated centre of the cluster, with circle radii defining the angular extent of the detected system. The left panel shows the un-merged catalogue, consisting of 1,623 clusters detected from all photometric filters, the right panel shows the effect of applying the merging prescription described in §3.3.6, selecting 102 clusters from the *left* panel.

discarded from the catalogue. However, to each cluster selected in this way, we attach a record of the candidate cluster galaxies that were not selected, forming an auxiliary catalogue of *associate cluster members*. In this way, we can keep track of galaxies the detector considered as members but did not include in the cluster. The degree of oversampling in colour space and hence number of multiple detections depends on the sampling interval dc, relative to the width $\sigma(c_{m20})$ of the filter. We set both of these parameters in §3.3.7.

3.3.7 Setting algorithm parameters

This section defines the values adopted for the algorithm parameters described in §3.3.2-§3.3.5.

Photometric filtering

In both mock and real datasets, we limit our search for clusters to three colours: *g-r*, *r-i* and *i-z*. These are used to form two sets of joint filters: {*g-r*, *r-i*} and {*r-i*, *i-z*}.

Each photometric filter is described by a colour normalisation c_{m20}, slope $\beta(c_{m20})$ and

#	Constraint
1	$(cg_{1,2}$ OR $cg_{2,1}) \geq 0.5$
2	$(pe_{1,2}$ OR $pe_{2,1}) > 0$ AND ΔS_1
3	BCG_{id} AND ΔS_2
4	$(pe_{1,2}$ OR $pe_{2,1}) \geq 0.8$ AND ΔS_3
5	$(pe_{1,2}$ OR $pe_{2,1}) \geq 0.8$ AND ΔS_4

Table 3.1: The set of conditions used to consider whether two clusters are multiple detections of the same system. If any one of these conditions are satisfied, the algorithm picks the "best" cluster of the two.

width $\sigma(c_{m20})$. The detection algorithm uses photometric filters that overlap in colour-magnitude space, preventing clusters close to filter edges from being poorly sampled. A sampling interval in colour space of $dc = 0.04$ is chosen, corresponding to an overlap of approximately 75% between adjacent filters based on σ_f, the filter width. For this study we demonstrate the detector with constant filter slope and width. In order to set these parameters, for each dataset, we identify a "reference cluster" from which we define the characteristics of the red sequence we are searching for.

SDSS reference cluster

To set β and σ for each colour in this dataset, we attempt to generate colour magnitude diagrams from the cluster members of Abell 2631 (Abell et al. 1989). With a measured redshift of $\mathrm{z} = 0.278$ (Crawford et al., 1995), this system is also the richest Abell cluster in Stripe 82. Ideally, one would produce colour-magnitude relations from spectroscopically confirmed cluster galaxies. Searching for redshifts in the NASA Extragalactic Database (NED[5]) between $0.2 \leq \mathrm{z} \leq 0.4$ and within 4' of the cluster centre, we find 5 galaxies with spectra, including the aforementioned BCG redshift. We identify these galaxies as red crosses in *g*, *r* and *i* composite Stripe 82 imaging of this cluster in Figure 3.9, and include $0.5r_{500}$ (Laganá et al., 2010)[6] as a dotted *grey* circle centred on the cluster BCG. The westernmost galaxy hosted Supernova SN2007pw, and was consequently

[5] `http://ned.ipac.caltech.edu`

[6] r_{500} is the radius within which the mean density is 500 times the critical density of the universe.

subject to spectroscopic followup at the Hobby-Eberly Telescope (Bassett et al., 2007). The blue disk galaxy (highest declination cross in the Figure) is unlikely to be a cluster member. The remaining galaxy measurements were all from the SDSS survey.

The low number of spectroscopically confirmed members in this cluster prevents the robust estimation of colour magnitude relations. As an alternative, we visually identify the cluster members from Figure 3.9, based on the colour and morphology of the galaxies. With this approach, we assign 126 galaxies to the Abell 2631 cluster, and show these putative members as blue circles in the Figure. Whilst the selection of these galaxies remains subjective, it permits a more robust estimate of the colour-magnitude relations (which show evidence of a clear sequence in all three colours used in this survey) and can be used to gauge, against our visual impressions, cluster purity in automated re-detections of this system.

A linear fit to the colour-magnitude sequence was applied to determine β for each colour. The filter widths were set using a method akin to that described in Gladders et al. (1998); we first remove the slope in each sequence and then exclude 3σ outliers. Starting at the line fitted to the cluster sequence, we increase the width in equal amounts above and below this line until we enclose 90% of the remaining members. We define this as the filter width σ for that colour.

Figure 3.10 shows the colour-magnitude sequence of the identified members in the three colours (*top*) compared to a field of the same area with no cluster present (*bottom*). *Blue* (*red*) points identify members that were inside (outside) the 3σ cut used to identify outliers. *Grey* data correspond to galaxies that were within $7'$ of the cluster centre and not picked as cluster members. Table 3.2 lists the fitted filter parameters for each colour (corresponding to the *black* lines in Figure 3.10) in addition to the colour range and number of filters used in our cluster search. Following our decision in §3.3.2 to use a fixed slope, we adopt the largest filter width (σ_f, 0.152) for all colours, Our chosen reference cluster should not bias our cluster detection from `ORCA` . Because we scan through colour-magnitude space in progressively redder selections (i.e. increasing $c_{\mathrm{m}20}$), we search for more distant systems. Although we fix the slope, our choice of photometric selection width encompasses a range of sequence gradients large enough to account for evolution as the algorithm searches to deeper redshifts. Analysis of mock clusters from the Millennium Simulation (Springel et al., 2005) suggests this approach probes at

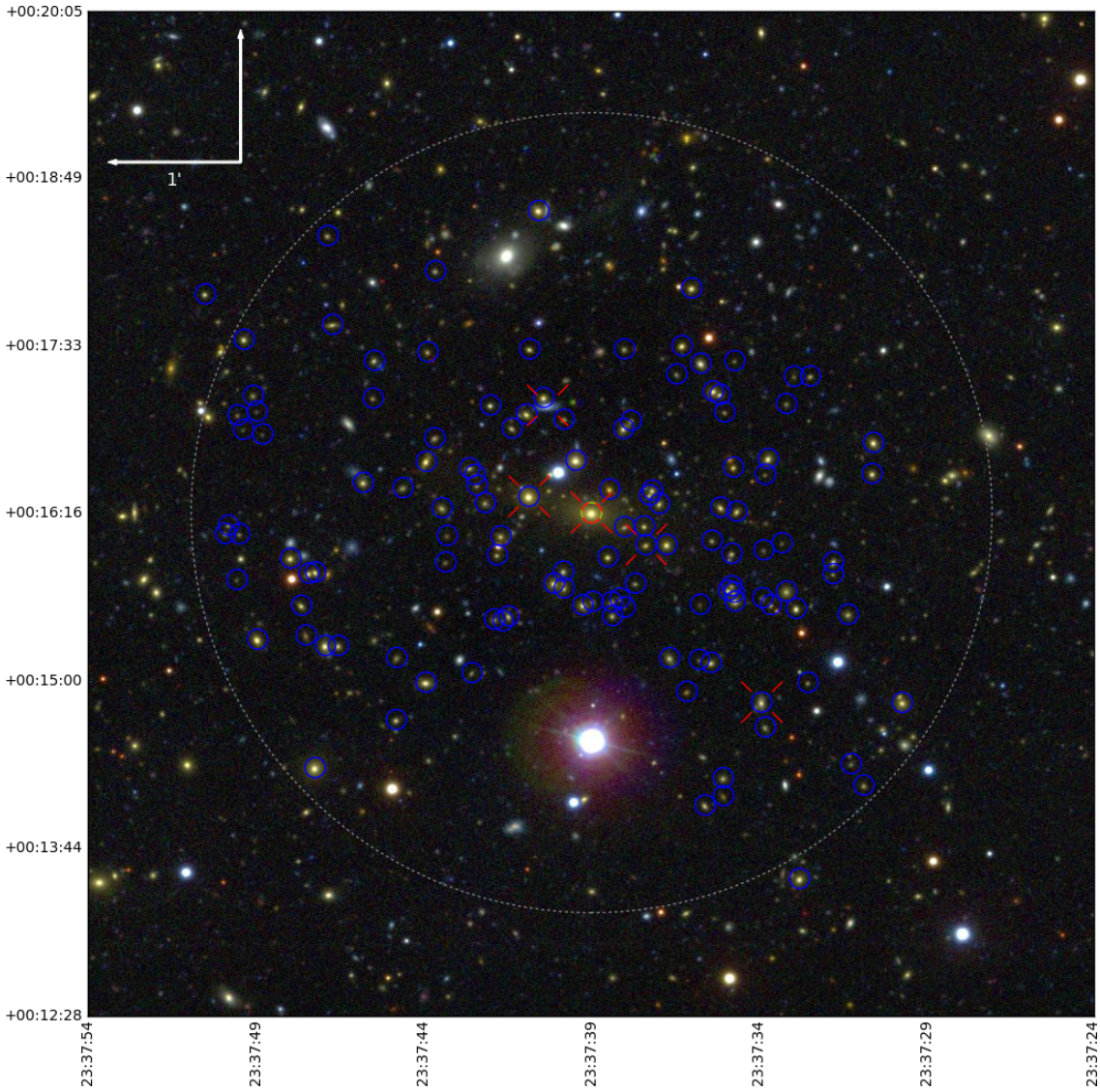

Figure 3.9: A *gri*-composite image of Abell 2631 from SDSS Stripe 82 imaging data. Small blue circles indicate the 126 galaxies visually selected as members of the cluster, which are used to fix the photometric filter slope and width. The yellow and red circle indicates the brightest galaxy selected. Galaxies with spectroscopic redshifts found in the NED are denoted by red crosses. For reference, we include $0.5\mathrm{r}_{500}$ as the grey dotted circle centred on the BCG.

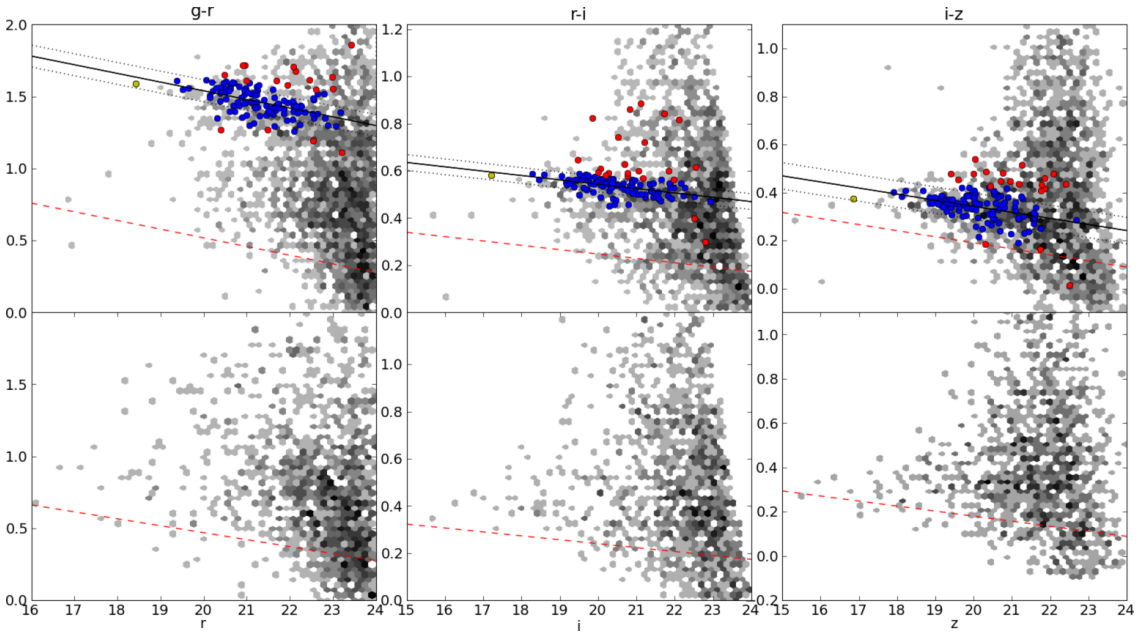

Figure 3.10: (*top*) Colour-magnitude diagrams for the 126 Abell 2631 members selected in this study. The yellow dot notes the position of the cluster *r*-band brightest cluster galaxy. The black lines denote filter fits to the data and indicate the slope (β), normalisation (solid, c_{m20}) and width (dotted, σ). The identified members are split into those inside (blue) and outside (red) the 3-sigma cut used to estimate the filter width. Grey data indicate all galaxies that were not identified as members of the cluster out to a radius of 7-arcminutes from the cluster centre. The red dashed line in the *g-r* colour indicates the blue limit imposed by the Virgo cluster, and the equivalent lines in *r-i* and *i-z* denote the lowest c_{m20} identified from cluster sequences in our search of the 7deg^2 Stripe 82 survey. (*bottom*) The colour-magnitude diagrams for galaxies in a region of the same area located in a field environment.

Colour	Slope (β)	Width (σ)	Range	1^{st}-stage filters
g-r	-0.048	$\mathbf{0.152}$	$0.47 - 2.00$	39
r-i	-0.017	0.067	$0.00 - 1.22$	38
i-z	-0.023	0.110	$-0.10 - 1.10$	31

Table 3.2: Filter parameters fitted from Abell 2631, the ranges searched and the number of filters in each colour. The blue limit in *g-r* corresponds to an extrapolation of the Virgo CMR, whilst the others permit a full sweep of the available data. The largest filter width (emboldened), defined as σ_f, is adopted for all colours.

least 2.5(1.5) magnitudes fainter(brighter) than the observed characteristic galaxy flux at the redshifts clusters are detected in this study. With measurements from a large ORCA cluster catalogue, future refinements to the algorithm may include a description of how the sequence slope varies with normalisation $c_{\rm m20}$.

The adopted filter width is used to define the input galaxy magnitude limit for each band. Magnitude limits are applied to reduce the number of input galaxies with high levels of photometric uncertainty. We define these as the faintest magnitude where the photometric uncertainties fall below $0.68\sigma_f$, setting limits in each band based on a sample of 100,000 galaxies from *Stripe 82*. Figure 3.11 shows the galaxy photometric error distribution for the *r*-band, and from this we set a magnitude limit of $r \leq 23.5$. This is slightly more conservative than the limit implied by the error distribution ($r \leq 23.8$) because we aim to include only sources with good photometry. The magnitude limits applied are 24.0, 23.5, 23.3, 21.6 in the *g*, *r*, *i* and *z* bands respectively, resulting in a source catalogue of 69,797 galaxies.

The bluest filter pair we detect in is *g-r*. To prevent the detection of spurious systems bluer than the $z = 0$ red-sequence in this colour we determine a blue limit by extrapolating the colour-magnitude relation (CMR) for Coma (Smith et al., 2009) and Virgo (Rines and Geller, 2008) to $r= 20$. The $c_{\rm m20}$ normalisation for Coma (Virgo) was estimated as 0.6 (0.47); we use the latter as the bluest filter possible in *g-r*. Similar limits are not applied to the other colours, but the normalisation below which no sequences were detected in *r-i* and *i-z* is described in §3.4.1. Figure 3.10 shows these limits as *red* dashed lines.

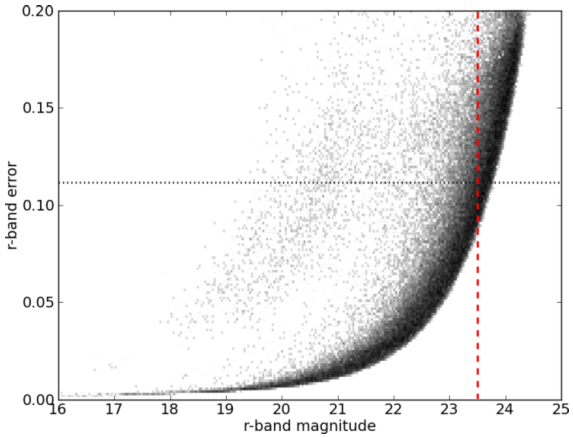

Figure 3.11: The SDSS model *r*-band photometric error in a sample of 100,000 Stripe 82 galaxies. These data are used to set a magnitude limit where at least 50% (0.68σ, black horizontal dotted line) of the faintest galaxies remain in a colour slice of width $\sigma_f = 0.152$. Whilst the data suggest a limit of $r \leq 23.8$, we opt for a slightly more conservative $r \leq 23.5$ limit (red vertical dashed line).

Mock survey reference cluster

We select a mock reference cluster from a set of ΛCDM-based detections generated from a preliminary scan of the simulation. The chosen cluster allows us to set the slope and width of the photometric filters in our search through the mock data. Candidate training clusters were identified from a redshift range bracketing Abell 2631 ($z = 0.278$), with similar memberships and a clear sequence in all colours. We selected the richest of these candidates, featuring 130 members and a redshift of $z = 0.3$. By applying the same fitting techniques as those described in §3.3.7, we set the filter parameters listed in Table 3.3 and apply the same colour ranges as those used on the SDSS. As seen in Figure 3.12, the fitted gradients are steeper in *g*-*r* and *r*-*i* than those used for the SDSS, and the filter widths are smaller. These values were nevertheless consistent with the other candidate reference clusters identified in the mock. As before, we use the most conservative width (*g*-*r*, 0.13) for filters in each colour.

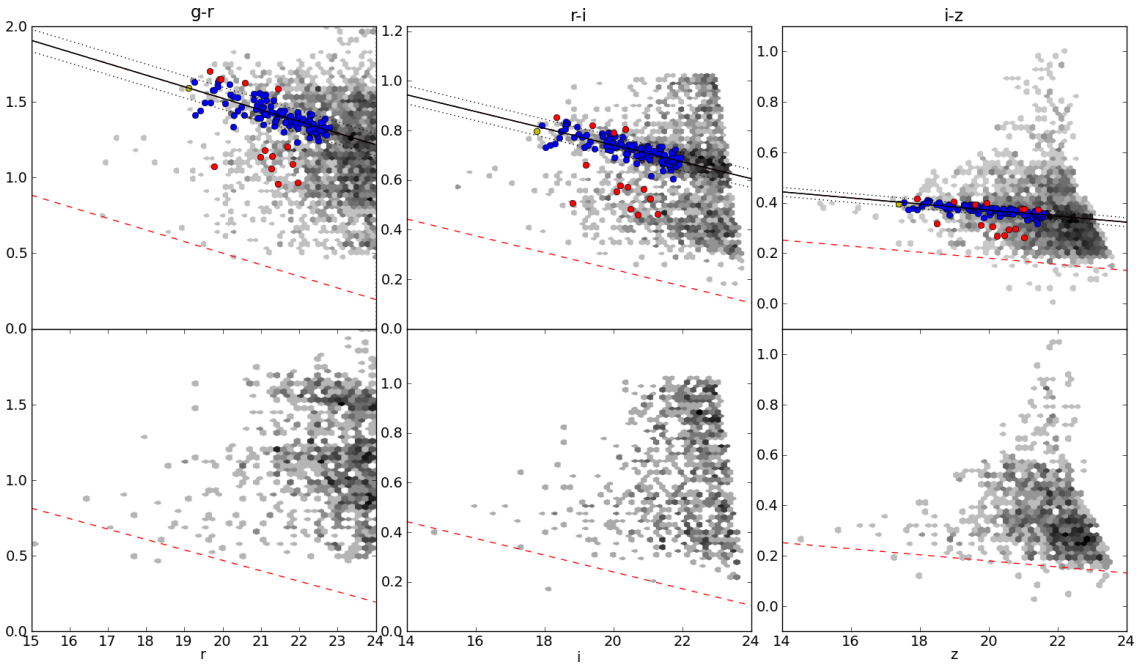

Figure 3.12: Mock CMR plots analogous to those in Figure 3.10. (*top*) Colour-magnitude diagrams for the mock reference cluster (*bottom*) The colour-magnitude diagrams for galaxies in the same area aperture, but located in a field environment. We have kept the same colour ranges used in the Abell 2631 data for comparison.

Colour	Slope (β)	Width (σ)	Range	1^{st}-stage filters
g-r	-0.065	$\mathbf{0.130}$	$0.47 - 2.00$	39
r-i	-0.032	0.064	$0.00 - 1.22$	38
i-z	-0.012	0.035	$-0.10 - 1.10$	31

Table 3.3: Filter parameters fitted from the mock reference cluster along with colour ranges searched by the detector (the same as those used in the SDSS data).

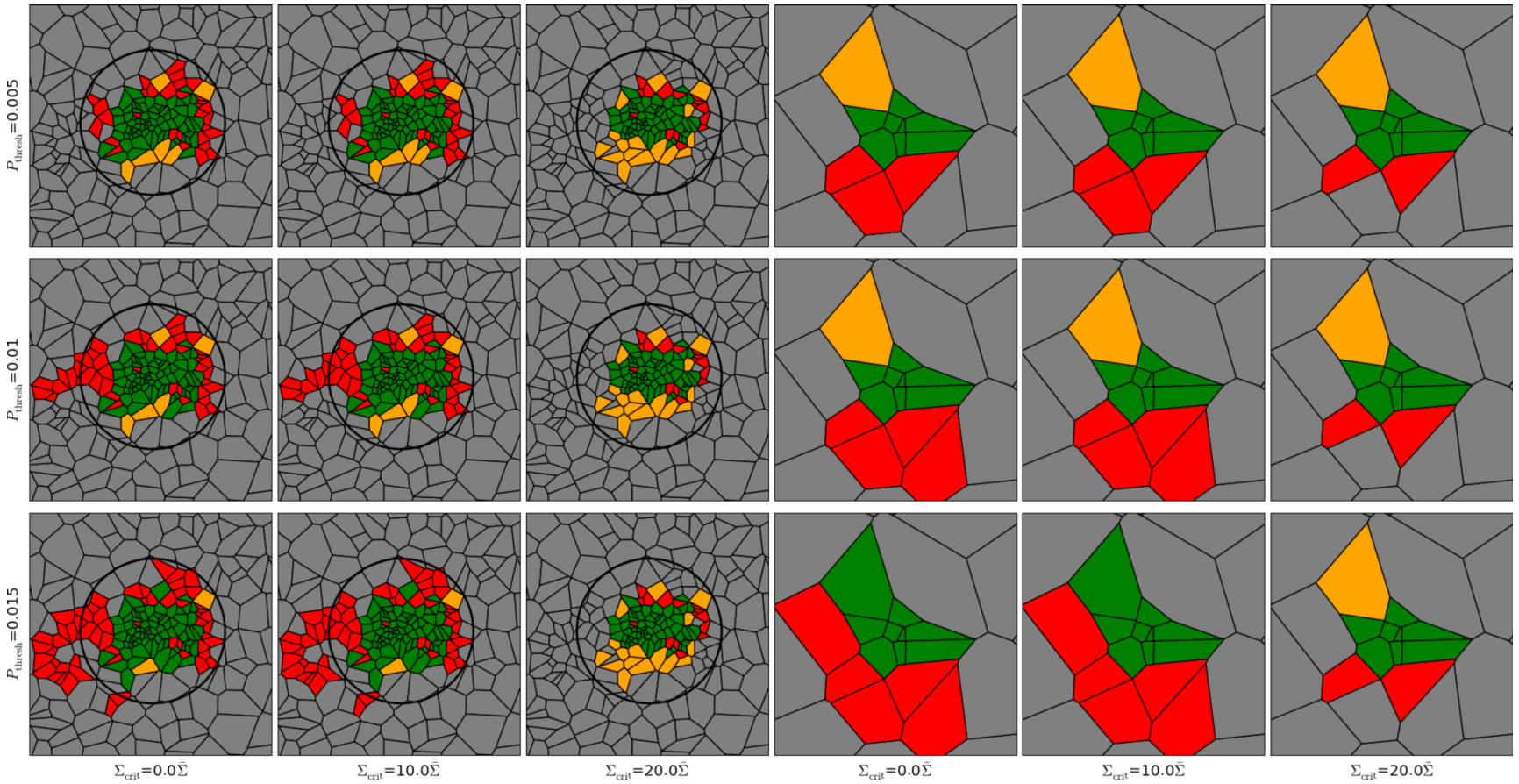

Figure 3.13: Effect of detection parameters on Abell 2631 (*left*, box scale 13.6′×13.6′) and a smaller group (*right*, box scale 3.5′×3.5′). Colour key: Grey are cells with field galaxies, green are galaxies identified by the algorithm that were also visually identified as members. Red cells are members assigned to the cluster by the detector but not visually identified as cluster members. Orange cells are galaxies that failed to be correctly identified by the algorithm as cluster galaxies, but were defined as such visually. The circle around Abell 2631 corresponds to a $1h^{-1}\mathrm{Mpc}$ radius at the cluster redshift.

Voronoi Tessellation and connection of overdense regions

The initial identification of clusters in projected high density regions and the subsequent percolation of their members depends respectively on the probability threshold $\mathrm{P_{thresh}}$ and the critical density Σ_{crit}. We parametrise the critical density Σ_{crit} as a scalar multiple of $\bar{\Sigma}$ such that both detection parameters have a mean density dependence. In the *left-hand* sequence of Figure 3.13, we note the effect a range of ($\mathrm{P_{thresh}}$,Σ_{crit}) combinations have on the recovery of Abell 2631 within a box of scale 13.6′. By tracking the detector's assignment of Voronoi cells to cluster and field, we compare members visually identified (see Figure 3.9) to the recovery of this cluster under different parameter combinations. The cells are colour-coded into four groups to differentiate detected and visually identified members. *Grey* cells show galaxies neither detected nor identified as cluster members. *Green* cells denote detected members that were also visually identified, *orange* for where the detector did not assign cluster membership despite our classification as such from the imaging, finally *red* cells are detected members not visually identified as members. We stress the latter group in no way indicates the purity of the cluster, as we are both incomplete and subjective in our identification of genuine cluster members. However, this exercise does provide a useful indication of detector performance when compared to our visual impression of cluster membership.

The detection grids show re-detection is broadly insensitive to the range of parameters explored. At higher probability thresholds (increasing row number) the cluster expands to form a more extended structure. This growth is moderated by the introduction of a minimum cell density. We exclude $\Sigma_{crit}= 20\bar{\Sigma}$ as it removes a significant fraction of visually identified members on the periphery of the cluster. The middle ground between detecting a more compact system ($\mathrm{P_{thresh}}$=0.005) and potentially increasing the interloper fraction ($\mathrm{P_{thresh}}$=0.015) suggests the balance of detection completeness and cluster purity lies with $\mathrm{P_{thresh}}$=0.01. We note from Figure 3.6 there are at minimum twice as many clustered cells as unclustered at $P(a) \leq 0.01$. Although (0.01,0$\bar{\Sigma}$) and (0.01,10$\bar{\Sigma}$) appear identical in their recovery of the cluster, we require a non-zero density constraint to filter out spurious low amplitude systems and prevent large clusters from percolating into giant connected structures. We consequently adopt the parameter combination ($\mathrm{P_{thresh}}$,Σ_{crit})=(0.01,10$\bar{\Sigma}$). To ensure these parameters are not biased to the detection of

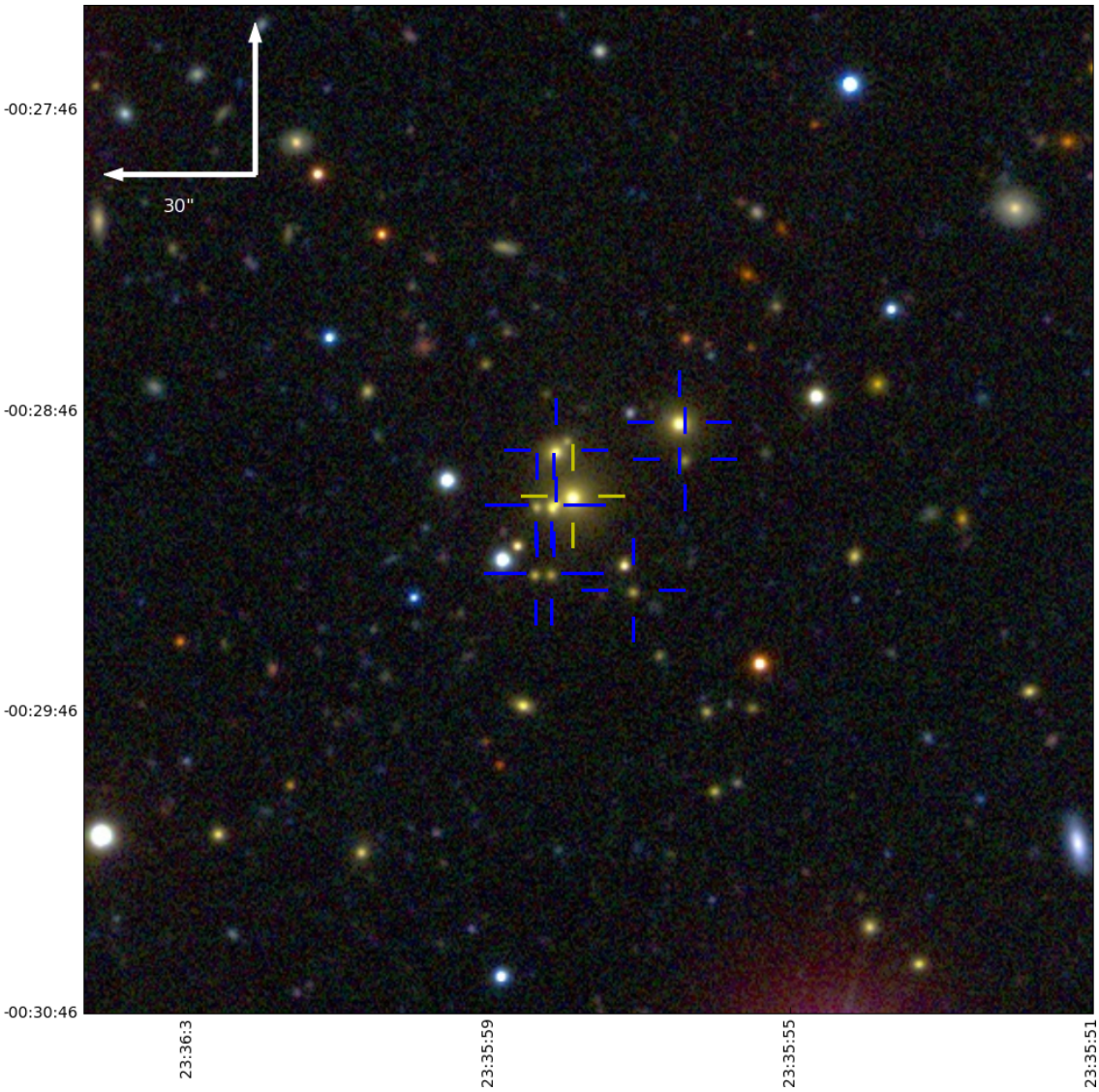

Figure 3.14: A *gri*-composite image of a group used to characterise how the ORCA recovery of galaxies in such a system relies on two detection parameters. Cross-hairs indicate visually-identified members, with the yellow cross-hair denoting the BCG.

high mass systems, we use 11 members of a visually identified group (shown in Figure 3.14) to perform a re-detection in the same parameter ranges. The *right-hand* sequence in Figure 3.13, with boxes of scale $3.5'$, shows the recovery of this group, and indicates group scale detection is robust to the range of parameters explored. The trade-off between completeness and purity is similarly evident here, with $(0.01, 10\bar{\Sigma})$ remaining a good compromise between the two.

In both cases (and more generally) there is a tendency to underestimate the total number of cluster members. This arises from an inevitable feature of Voronoi Diagrams implying the algorithm is unlikely to recover all cluster members. The suppression of the field galaxy population with photometric filters causes an abrupt drop in galaxy surface density at the cluster boundary. Because the Voronoi cells of peripheral members have a limited number of field galaxies to constrain their boundaries they adopt larger areas. Such cells may then be rejected as members because their areas are inconsistent with that population. Nevertheless, tests with mock catalogues allow us to quantify the impact this effect has on the cluster purity, as discussed later in §3.6.

Finally, we set the minimum membership of a cluster, N_{min}, to five galaxies.

3.4 SDSS Equatorial stripe 82 sample cluster catalogue

In this section, we describe the application of the algorithm to real astronomical data.

3.4.1 The catalogue

We applied the detector to a 7 square degree sample of Stripe 82, using the limits described in §3.2 and parameters described in §3.3.7. Here we describe the general characteristics of this catalogue, perform a series of tests on the data and briefly compare our detections to existing optical and X-ray-detected clusters.

After applying the magnitude limits described in §3.3.7, a source catalogue of 69,797 galaxies is analysed by the algorithm. We find a total of 97 clusters, identifying a total of 1293 cluster galaxies (0.5% of the original galaxy sample) and 813 associate cluster members (candidate cluster members that were not selected). Of these clusters, 34% were detected in $\{g\text{-}r, r\text{-}i\}$ and 66% in the $\{r\text{-}i, i\text{-}z\}$ combinations.

Although we define a blue limit for the *g-r* colour-magnitude relation ($c_{\mathrm{m20}} > 0.47$), equivalent limits were not applied to the *r-i* and *i-z* colours. We can however place upper bounds on the blue limit in these colours by noting no clusters were detected below *r-i*=0.24 and *i-z*=0.18. Such limits serve to reduce the search time for future survey scans.

Table 3.4 shows an extract of the cluster catalogue, the full dataset can be found in §C.2 of Appendix C. This 7 square degree sample of 97 Stripe 82 clusters is also avail-

able online[7]. Each cluster is named according to the IAU convention, in the form MGB JHHMMSS+DDMM.m. We detail below the main features both catalogues.

3.4.2 Cluster positions & redshifts (`cluster_z`, `cz_type`)

The `ra` and `dec` position quoted in the catalogue is the algorithm estimate of the centre of each cluster, based on the average positions of their members.

Although we do not use any redshift data to generate our cluster catalogue, we provide redshift estimates for each system detected by the algorithm. These redshifts are weighted towards members with spectroscopic data, but two sets of photometric redshift data (*hyperz* and the DR7 photometric estimate) are used to provide each cluster galaxy with at least one redshift estimate. From the catalogue of 1293 cluster galaxies, 2.6% have spectroscopic data (DR7 spectroscopic redshifts, WiggleZ and 2SLAQ), 93% have DR7 photoz and 87% have `hyperz` estimates. The `hyperz` estimates for cluster members were generated using only S0 and E SEDs, a Calzetti et al. (2000) reddening law and a two-stage convergence (over and above that performed by `hyperz`) to the redshift where a range identified in coarse redshift bins is re-sampled with a smaller bin width. As seen in §3.2.2, the error dispersions against spectroscopic redshifts are higher for the *hyperz* data than for the DR7 photoz (0.029 vs 0.016).

We calculate each cluster redshift by determining the weighted median redshift from the available member data. The weighting for a spectroscopic, DR7 photoz and `hyperz` redshift is 4, 2, 1 respectively, the higher weighting for DR7 photoz reflecting the smaller error dispersion mentioned above. To gauge the accuracy of our redshift estimate, we note the calculated redshift of Abell 2631 is $\mathrm{z} = 0.26$, some 0.02 lower than the value determined by Böhringer et al. (2000). The median cluster redshift of the whole catalogue is $\mathrm{z_{med}} = 0.31$, and the maximum redshift is $\mathrm{z} = 0.57$. Approximately 25% of the clusters have at least one member with a spectroscopically measured redshift.

The `cz_type` property is a shorthand description of the available redshift data for each cluster, each letter defining a measurement type, followed by the number of that type. The letters denote data from the mo(c)k, DR7 (s)pectroscopic, (w)iggleZ, 2SLA(q), DR7 (p)hotometric and (h)yperz datasets, where mock is of course not used in this ob-

[7] `http://astro.dur.ac.uk/~dmurphy/orca`

Name	RA	dec	cluster_z	cz_type	N_{gal}	b_gc	scatter	θ_{80}	θ_{80}/θ_{20}
MGB J234017-00030.9	355.06912	-0.06455	0.245	c0s0w0q0p6h2	6	2863	0.088	0.007	1.700
MGB J233817+00190.0	354.56897	0.33309	0.208	c0s0w0q0p8h6	8	27748	0.077	0.018	3.667
MGB J234113-00000.4	355.30349	-0.00597	0.166	c0s0w0q0p6h2	6	15098	0.035	0.016	1.692
MGB J234400-00300.3	355.99952	-0.50461	0.181	c0s1w0q0p5h4	6	4477	0.001	0.008	1.750
MGB J234725+00190.7	356.85322	0.32867	0.201	c0s0w1q0p14h14	14	9562	0.069	0.020	2.545

Table 3.4: A sample of the ORCA cluster catalogue generated in this study. Full details of the columns can be found in §3.4.1-§3.4.5. The first column contains the cluster name based on the IAU convention. Columns 2 and 3 note the J2000 estimated cluster positions in degrees. Columns 4 and 5 describe the cluster redshift and source data used to calculate the redshift. Column 6 notes how many members were found in the cluster, and we provide estimates for the cluster $B_{\rm gc}$ richness and sequence scatter in Columns 7 and 8. The final two columns indicate the radius (in degrees) enclosing 80% of the cluster members and the ratio of this value to the 20% radius, a measure of cluster concentration.

servational data.

3.4.3 Cluster richness (b_gc)

With access to cluster redshifts we are able to calculate the B_{gc} optical cluster richness, a robust parameter known to correlate with cluster mass. We use the B_{gc} measure described in Yee and López-Cruz (1999):

$$B_{gc} = \frac{\rho_{bg} D(\mathrm{z_{cl}})^{\gamma-3} A_{gc}}{I_\gamma \Psi[M_3, M_3+3, \mathrm{z_{cl}}]} \tag{3.3}$$

where ρ_{bg} is the background surface density of all source catalogue galaxies inside a $0.5h^{-1}\mathrm{Mpc}$ radius with luminosities between the third brightest cluster galaxy (M_3) and three magnitudes fainter. D, the angular diameter distance, is derived from the cluster redshift $\mathrm{z_{cl}}$. γ and I_γ respectively define the slope of the angular galaxy correlation function and the integration constant arising from de-projecting the cluster, and are set to $\gamma = 1.77$ and $I_\gamma = 3.78$. The correlation amplitude A_{gc} is defined as:

$$A_{gc} = \frac{N_{net}}{N_{bg}} \frac{(3-\gamma)}{2} \theta^{\gamma-1} \tag{3.4}$$

where N_{net} is the background-corrected count of galaxies within the luminosity range described above, out to an angular separation θ that corresponds to $0.5h^{-1}\mathrm{Mpc}$ at the cluster redshift. N_{bg} is the background galaxy count within this radius, estimated from the mean density of galaxies across the whole field. We provide a more detailed appraisal of cluster richness estimates in chapter 4. Later in §5.5, we attempt to calibrate cluster mass estimates to optical richnesses.

3.4.4 Cluster sequence scatter (scatter)

To estimate the width of a detected cluster's sequence, we first make a fit to the slope of the sequence and remove the tilt. Using cluster members between $m_{\mathrm{BCG}} \leq m \leq m_{\mathrm{BCG}}+3$, we estimate the sequence scatter by making a 2σ clip in the colour distribution.

3.4.5 Projected scale (θ_{80}, θ_{max}) & concentration (C)

For each cluster, a projected scale size θ_{80} is provided. This is calculated as the angular radius (in degrees) enclosing 80% of cluster members from the centre.

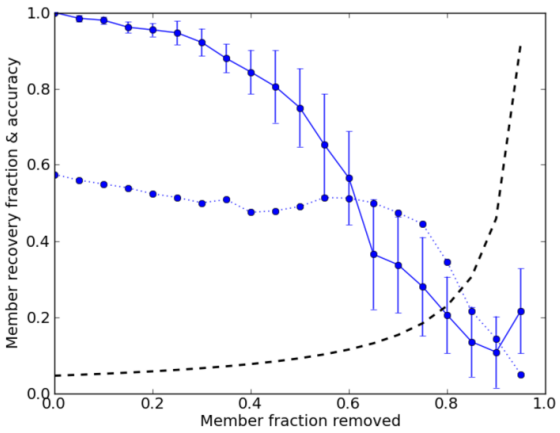

Figure 3.15: The recovery fraction (solid line) and recovery accuracy (dotted line). Some Abell 2631 cluster galaxies are randomly removed from the source catalogue, and the fraction subsequently identified in a re-detection of the cluster is the recovery fraction, with error bars of 1σ uncertainty calculated from 50 re-detections. The fraction of visually identified Abell 2631 galaxies making up the re-detected cluster defines the recovery accuracy. The fraction required to produce an N_{min}=5 member system is denoted by the black dashed line.

A measure of the projected concentration (C) is determined by comparing the radius enclosing 80% of the cluster members to the radius enclosing 20%. High values of θ_{80}/θ_{20} indicate a centrally concentrated cluster.

3.4.6 Testing the algorithm

Cluster re-detection robustness

To determine how robust the detector is to catalogue incompleteness, we attempt re-detections of the Abell 2631 cluster after removing a random selection of members from the source data. Our sole constraint is that the cluster BCG remains in the source data. In the following analysis, we only consider the detected cluster closest to the original Abell 2631 position. Robustness is defined as the fraction of members detected in the new cluster from those remaining in the input catalogue. We use a test *g-r* photometric filter that adopts a β_{g-r}, $c_{\rm m20}$ and σ_{g-r} best suited to the recovery of A2631, selecting approximately 85% (108) of the visually selected members. We experiment with removal

fractions down to 95%, corresponding to the largest fraction still retaining $N_{min} = 5$ original cluster members in the sample.

Fifty random realisations of a depleted input catalogue are generated for each removal fraction, yielding a median recovery rate based on members that could have been added to the cluster. The solid blue line in Figure 3.15 shows how increasing the removal fraction affects the fraction of cluster members recovered; error bars on this line represent 1σ uncertainties from the 50 re-detections in each bin. The recovery fraction when no galaxies have been ejected is $\sim 93\%$ of the 108 A2631 members located inside the photometric filter. The other 7% were rejected by the algorithm because either their Voronoi cells have insufficient densities to join the overdense collection of cells ($\mathrm{P_{thresh}}$, see §9), or their inclusion causes the percolating cluster to drop below the critical density (Σ_{crit}).

We take into account this intrinsic detection inefficiency, quoting yields from the cluster re-detection relative to the $\sim 93\%$ of members recovered where no additional galaxies are removed. Unsurprisingly, the fraction of detected members located in the cluster drops as more members are excised. However, over 75% of remaining members are re-detected even after half of the cluster is removed. Approaching larger removal fractions, the fragmentation of cluster members into spatially distinct groups hinders recovery of the complete set. The black dashed line in this plot corresponds to the minimum recovery fraction required to identify $N_{min} = 5$ original members from the input data. The algorithm can robustly identify the original cluster down to an 80% removal fraction, corresponding to 22 of the original 108 galaxies. Below this limit, an insufficient number of cluster members are recovered by the detector to identify a cluster associated with the halo.

For each ejection fraction we also calculate the recovery accuracy: the fraction of visually identified A2631 galaxies making up the re-detected cluster. The dotted blue line in Figure 3.15 shows this parameter. The initial accuracy (no members are removed) is approximately 60%, providing some estimate of our level of incompleteness when visually identifying cluster membership[8]. As more members are removed, there is a gradual reduction in accuracy, implying replacement of these members with other galaxies

[8]This can be seen in the (0.01,10$\bar{\Sigma}$) panel of Figure 3.13 by comparing the ratio of *red* cells to the number of *green* and *red*

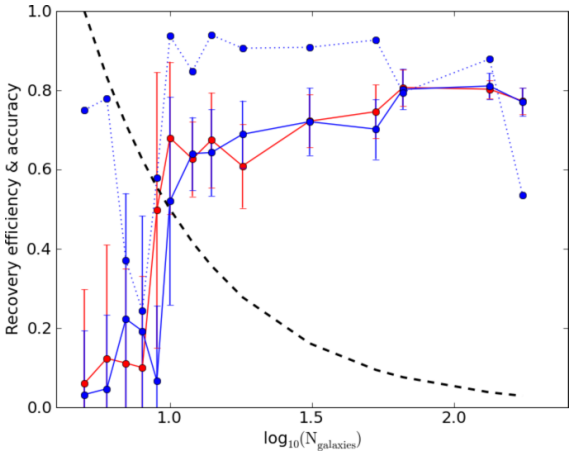

Figure 3.16: The algorithm's re-detection capability when a cluster has been moved to a random position. The recovery efficiency (solid blue line) is the fraction of original cluster galaxies found in the displaced cluster. The edge-effect recovery efficiency (red line) shows a similar test, instead moving the cluster to a random position near the survey boundary. Uncertainties in both lines are 1σ errors from 50 re-detections. The recovery accuracy (dotted blue line) is the ratio of input cluster members to the member count of the re-detected cluster. The black dashed line indicates the N_{min}=5 threshold required to secure a robust detection of the cluster's halo.

becomes more commonplace. At large ($> 70\%$) removal fractions, fragmentation acts to reduce the connectivity of cluster members, increasing the number of contaminant galaxies that share the photometric filter.

Cluster displacement and edge effects

A cluster detector should identify systems irrespective of the environment they are located in. Ideally then, recovery of identified members is achieved even if the system is moved to another position.

To determine the sensitivity of cluster identification to localised background fluctuations, we shift source data positions of known cluster members to a random location, keeping their spatial distribution intact. A buffer is created around the survey edge to ensure no cluster members are displaced outside the boundaries, then a re-detection of

the cluster is attempted. The re-detection performance is quantified by the recovery efficiency - the fraction of original members in the new cluster, and the recovery accuracy remains as defined in the previous test.

Figure 3.16 shows the recovery efficiency (*solid blue*) and recovery accuracy (*dotted blue*) for clusters spanning more than an order of magnitude in membership ($N_{min} = 5$ to 174 galaxies). If there was a choice of cluster for a membership bin, we used the system with the smallest sequence scatter to determine the impact of displacement on the best candidate in that membership group. Each cluster was re-detected in the pair of filters it was originally identified in, meaning a re-detection with no displacement would yield a perfect recovery efficiency and *recovery accuracy* (both equal to one). We perform 50 random displacements for each of the selected clusters, using their scatter to derive 1σ uncertainties from the mean. The *black dashed* line corresponds to the recovery fraction required to detect $N_{min} = 5$ galaxies of the original system from each displaced cluster.

For the majority of cluster sizes, recovery accuracies are approximately constant at $\sim 90\%$, meaning 10% of the cluster members are background galaxies selected in the same filter. Recovery efficiency data suggest the detector makes significant cluster re-detections for systems down to 10 members, but smaller groups are susceptible to higher levels of contamination and fragmentation. Our example case of Abell 2631 (at $log_{10}N_{gal} \sim 2.1$), with a recovery efficiency of 80% is approximately 13% lower than the recovery fraction from robustness test calculated above. A recovery accuracy of ∼86% is consistent with the detector swapping 13% of original members with background galaxies when the cluster is moved.

We next establish how survey edges bias the detection of systems at the boundaries. Using the same set of clusters, we repeat the above experiment, specifically placing systems close to the survey edges to quantify the impact of edge effects on group and cluster recovery.

Galaxy cells at the boundary of a Voronoi Diagram are unbounded, often resulting in very large cell areas. This may hamper the identification of low-membership clusters, where a member with cell area exceeding the probability threshold may preclude the cluster from detection. Random positions are selected along any one of the four sides of the survey (allowing clusters to reside in a corner). In our source catalogue, the declination boundaries (at $\delta = \pm 1.25°$) are set by the geometry of the stripe, whilst the RA

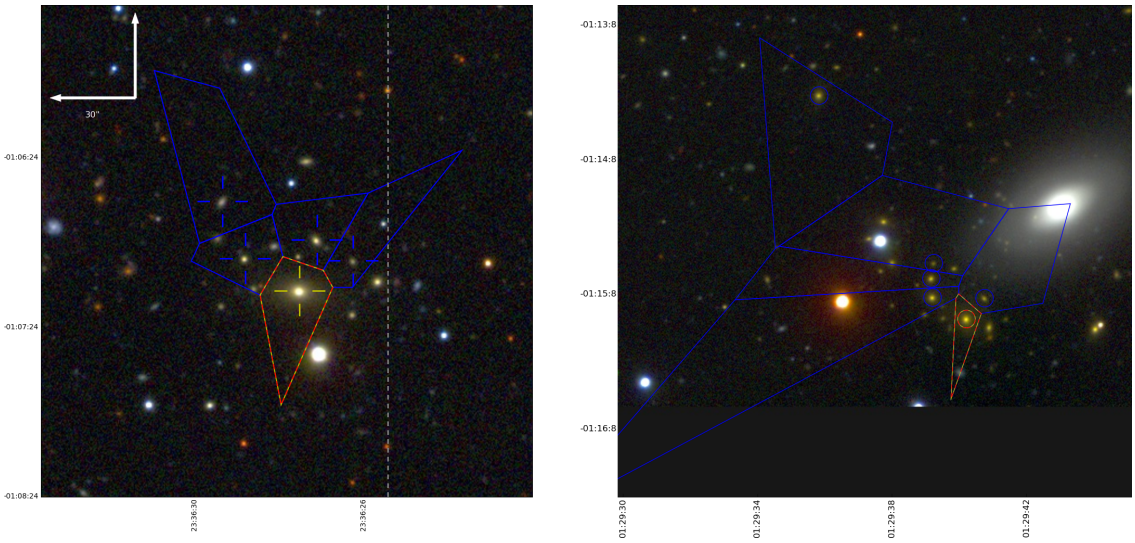

Figure 3.17: Two group-scale systems detected at the very edge of the Stripe 82 survey limits. The *left-hand* cluster has a limit enforced in RA by the data range ORCA explored (grey dotted line), whilst the *right-hand* cluster is limited by the actual survey edge.

boundaries are artificially defined. Distances between the cluster centroid and survey edge are large enough to include all members within the survey. The red line in Figure 3.16 shows the recovery efficiency based again on 50 randomised displacements. This distribution is very similar to that of the displacement test above, suggesting edge effects do not hinder the recovery of clusters any more than the displacement of the members themselves. This is particularly significant at group scales, where the exclusion of one or two members could prevent the detection of the system. In Figure 3.17 we show two examples of ORCA groups detected close to the edge of the survey. The boundary in the *left*-hand group was set artificially by the algorithm, whilst the *right*-hand boundary is the actual declination edge of the stripe.

False positive detection rate

We set the detector the task of attempting to detect spatially clustered systems with randomised colours. This establishes the importance of red sequences to cluster detection with this algorithm and provides an estimate of the false detection rate. We run the detector on the source catalogue in the same manner as before, having first shuffled the colours so while cluster members still reside in high surface density regions, they no

longer have red-sequences. We identified two "clusters" (with 5 and 6 members) in the 7 square degree survey, both located at the positions of original high-membership ORCA clusters. To ensure this calculation is uninfluenced by the size of the survey, we repeat this process on the full Stripe 82 dataset ($-50^\circ < \alpha < 59^\circ$) covering 272.5 square degrees. The algorithm detects 15 "clusters" from these data, each consisting of five or six-member groups. From this we infer the number of spurious systems detected per 7 square degrees is 0.39.

In a similar fashion we next randomise galaxy positions while keeping the colours the same. This means cluster red-sequences remain intact as the algorithm scans through colour-magnitude space, but points clustered in colour are no longer clustered in the sky. The algorithm detected four "clusters" over the full 272.5 square degree Stripe 82 dataset, implying a $\sim$0.1% spurious cluster detection rate.

Both exercises suggest the detector cannot identify clusters without correlations in both colour and spatial position. Moreover the probability of detecting systems based on random distributions of both colour and position is below 1%.

Because this algorithm is designed to identify correlations in colour and spatial position, where systems are projected neighbours and share similar colours, it remains possible the ORCA detector will identify them as a cluster. It is also possible that such detections persist over multiple filter combinations. This can be seen in Figure 3.18, where we show ORCA detections of a group of (presumably) unassociated sources. These have been identified in three separate filter combinations, and one would be selected as the "best" cluster. The redshift estimates of these groups, based on the photometric redshifts of their galaxies exhibit a large scatter. By design however, some also share similar redshifts, as they were selected in the same photometric filters. Combining the redshift data of all assigned galaxies, we find a median redshift of $z = 0.08$ comparable to the spectroscopic redshift of the large spiral (z_s=0.079) but with a 1-σ scatter of 0.11. Although we cannot conclusively rule out such detections, their members are atypical of the traditional cluster population.

One strategy to reduce the incidence of such associations is to minimise stellar contamination of the input catalogue. Where constituents appear to be genuinely extended sources (as appears the case here), one could look to adding more constraints to the observed properties of the sequence - one example would be filtering out systems where

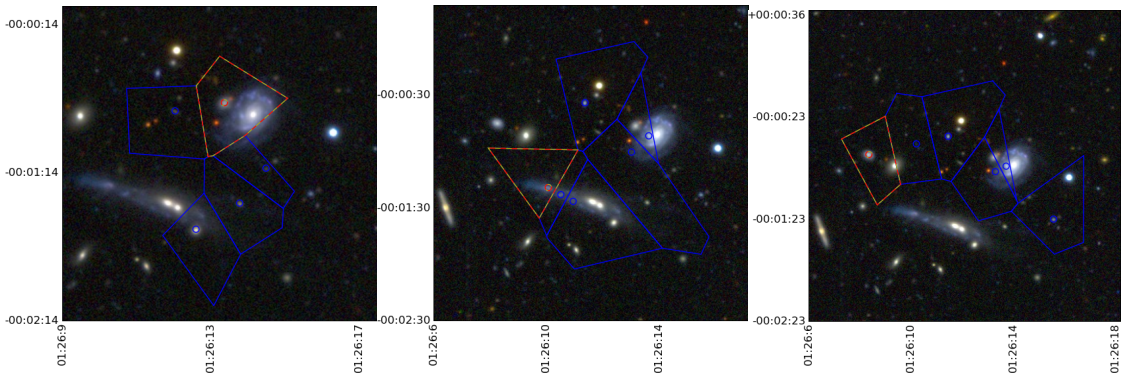

Figure 3.18: Example of a spurious ORCA detection, arising from chance associations of similarly coloured galaxies. Because this conglomeration appears in multiple selection filters, (as seen in these three panels) ORCA detects it more than once.

there is a significant separation between fluxes of the putative BCG and the next brightest member. Moreover, the algorithm may also detect clusters from associations of "galaxies" that are actually the fragmented haloes around bright stars, uniformly (and highly) reddened galaxies in regions of high Galactic extinction and associations of "galaxies" from overly de-blended galaxies in the catalogue. Whilst we have taken care to reduce the incidence of these objects in our source catalogue, some degree of contamination will inevitably remain.

Projected cluster-pair resolution

The ideal algorithm can identify two clusters with the same angular position on the sky, but at different radial distances. Using the $c_{\mathrm{m20}} - \mathrm{z}$ relation demonstrated in Figure 3.5, one can in principle isolate superimposed systems by identifying them in different filters. Within a detection filter $f(C_{\mathrm{A}})$ of width σ_f, two spatially coincident systems will be merged even if their sequences do not directly overlap. We overcome this limitation by splitting sequences in the following colour (C_{B}) with the application of joint filters (§3.3.3). The resolving power of the algorithm in projection is therefore limited by the merging of separate clusters that are mistaken as multiple detections in §3.3.6.

We test this effect with the same clusters used in §3.4.6 by implanting a 7-member test cluster at the same spatial position and colour normalisation c_{m20}. We increase the

Cluster name	RA	dec	z_{photo}	z_{spec}	N_{gal}	N_{gal}^{R200}
BCG J233740+00160.3	354.41553	0.27138	0.286	0.277	59	88
BCG J234624+00440.0	356.59955	0.74943	0.273	0.275	25	26
BCG J233746-00420.2	354.44067	−0.70310	0.286	0.287	20	17
BCG J234100+00040.9	355.24905	0.08161	0.194	0.185	23	23
BCG J233955-00250.0	354.97916	−0.43282	0.275	0.277	17	15
BCG J234548-01070.7	356.45068	−1.12775	0.273	–	18	18
BCG J234604-00100.0	356.51477	−0.18283	0.254	–	22	22
BCG J234322+00190.6	355.84039	0.32587	0.257	0.267	38	60
BCG J234146+01070.5	355.44077	1.12444	0.246	0.251	15	11
BCG J233919-00150.6	354.82941	−0.25941	0.284	–	14	11
BCG J234024-00050.6	355.10205	−0.09300	0.281	–	17	13
BCG J234720+00290.7	356.83487	0.49456	0.286	0.275	12	10
BCG J233900+00420.0	354.75143	0.71610	0.219	0.183	14	11
BCG J234122+00190.0	355.34253	0.33330	0.284	0.278	22	22
BCG J233911-01130.3	354.79459	−1.22236	0.292	–	14	10
BCG J234626+00430.7	356.60690	0.72794	0.251	–	25	29
BCG J234403+00130.6	356.01273	0.22646	0.262	–	16	11
BCG J234233-00170.3	355.63776	−0.28873	0.275	–	16	14
BCG J233755+00130.5	354.47760	0.22478	0.262	0.278	37	61
BCG J233825-00090.2	354.60291	−0.15397	0.270	–	14	11
BCG J234737-00370.9	356.90375	−0.63221	0.262	–	14	11
BCG J234106+00120.4	355.27640	0.20707	0.262	–	15	10

Table 3.5: An extract from the Koester et al. (2007a) catalogue noting the 22 `maxBCG` clusters within the limits of this SDSS sample field. The cluster name follows the IAU JHHMMSS+DDMM.m format. The RA and dec are J2000, and measured in degrees. z_{photo} and z_{spec} are the estimated photometric and spectroscopic redshifts of the clusters. N_{gal} is the number of members in the cluster, and N_{gal}^{R200} is the scaled richness.

test cluster $C_{\rm B}$ colour normalisation by $\delta c_{\rm m20}$ and run the matching algorithm[9]. This is repeated until the detector classifies the reddened test cluster as an independent system. The *resolving capability* of the algorithm can be parametrised as $\chi = \Delta c_{\rm m20}/\sigma$: the minimum sequence colour separation between the two detected systems relative to the width of the filters they were identified in. Small values indicate a good resolution, and in all clusters tested against, we found $\chi < 0.5$. Moreover, for all but two membership bins (N_{gal}=14,18) the test cluster was resolved within $\chi < 0.25$. Whilst in our real astronomical data we observe some cluster pairs overlapping in projected space, these examples exhibit large separations in both colour space and redshift. For example the two clusters *MGB J234729-00080.4* and *MGB J234733-00100.0* have redshifts of $\rm z = 0.23$ and $\rm z = 0.53$ and $\chi_{\rm r-i} = 7.8$. Although our analysis here could benefit from a larger sample size, `ORCA` can distinguish between two separate systems even if their sequences lie in the same filter, subject to their colour separation being at least $1/4$ the filter width. Below this level, their similarity in colour likely justifies classifying these systems as the same structure. These cases will only be disentangled by follow-up spectroscopy which can determine the relative velocity offsets between potentially separate or merging systems.

3.5 Comparison with existing cluster data

The positions of detected clusters can be seen in Figure 3.19, with the location of `maxBCG` clusters (Koester et al., 2007a) marked with red circles, and the positions of known X-ray clusters marked with blue squares. Clusters detected in the {*g-r*, *r-i*} combination are shown as blue filled cells, those detected in {*r-i*, *i-z*} are red filled cells. In each case the cluster BCG cells are yellow.

3.5.1 The SDSS `maxBCG` catalogue

The Koester et al. (2007a) `maxBCG` catalogue of 13,823 optically selected SDSS clusters uses the detection algorithm described in Koester et al. (2007b). This catalogue makes use of data from an earlier release of SDSS, so was unable to take advantage of the added depth Stripe 82 offered this study. Because direct comparison of the two cluster selection

[9] Although this to some degree simulates a migration of the cluster to higher redshifts, we make no adjustment to the sequence slope or scatter - these will also impact to the resolving capability of the algorithm.

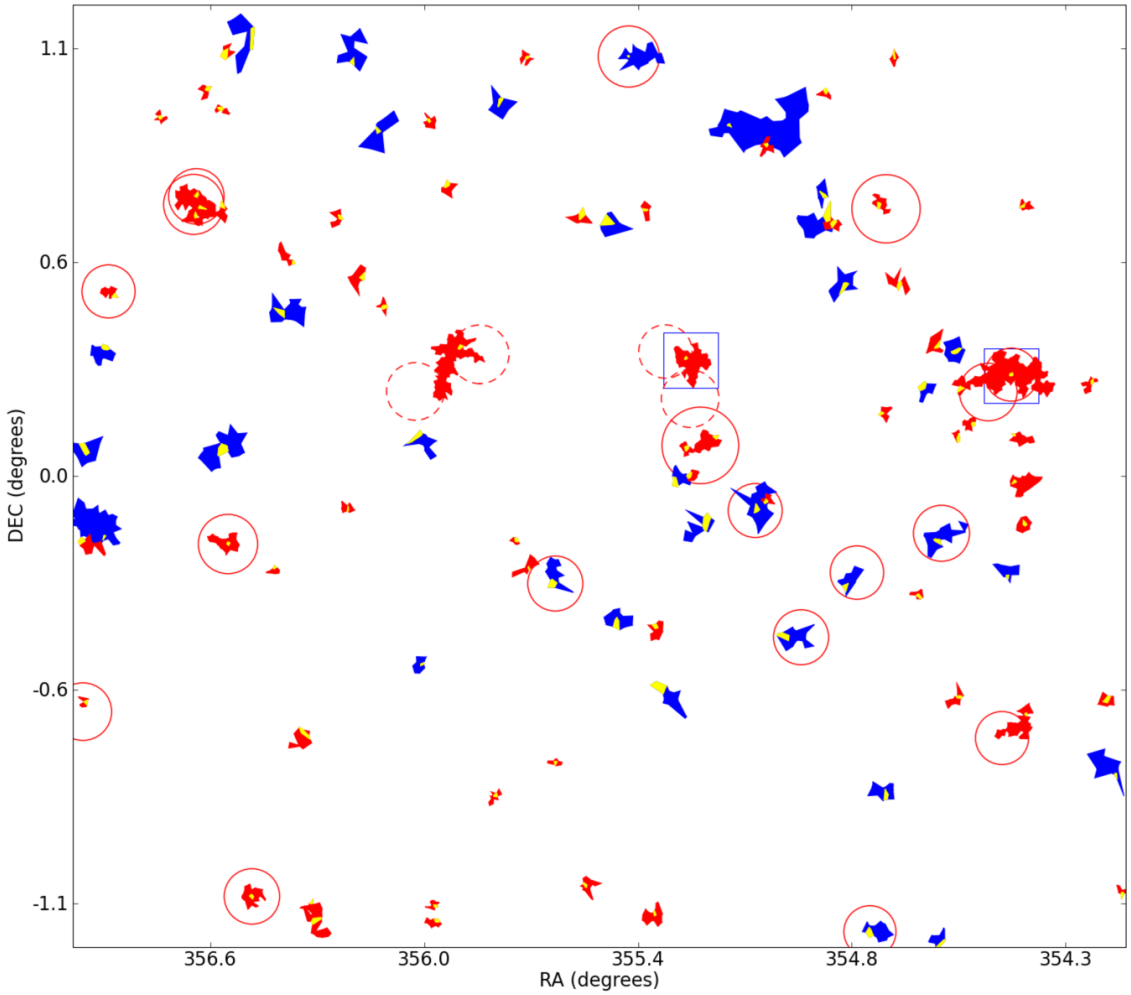

Figure 3.19: Clusters detected in the Stripe 82 field. The coloured cells represent clusters detected in different colour pairs. Blue cells correspond to clusters detected in $\{g\text{-}r, r\text{-}i\}$ filter pairs, red clusters detected in $\{r\text{-}i, i\text{-}z\}$ filter pairs. Yellow cells indicate the BCG position of each cluster. Red circles indicate the position of `maxBCG` clusters, based on data shallower than that used in the study here. Circle radii correspond to $1h^{-1}\mathrm{Mpc}$, based on the `maxBCG` photometric redshift estimate of the cluster. Dashed red circles indicate the four `maxBCG` clusters discussed in §3.5.1 that also feature *gri*-colour imaging in Figures 3.20 and 3.21. Blue squares note the position of X-ray clusters, with half-lengths corresponding to $1h^{-1}\mathrm{Mpc}$.

functions is both non-trivial and unfair, we do not attempt a full analysis in this study. However, in the spirit of matching detections made here to those of the shallower data in the Koester et al. (2007a) catalogue, we include the positions of maxBCG clusters in Figure 3.19 as a set of red circles. The centre of these circles is the location of the assigned Brightest Cluster Galaxy (BCG), whilst the radius corresponds to $1h^{-1}\mathrm{Mpc}$ calculated from the published photometric redshift estimate of the cluster. We stress however, that this does not necessarily correspond to the physical size of the cluster.

The survey area contains 22 maxBCG clusters. For ease of reference, salient details from that catalogue are reproduced in Table 3.5, along with a name of the form BCG JHHMMSS+DDMM.m. We attempt a simple match to the ORCA catalogue by looking for either common BCGs (and more generally a match to ORCA cluster members where BCGs are assigned differently) or statistically significant separations between ORCA centroids and maxBCG positions. We group our cluster detections into five categories:

1 *BCG match*: in this category, the position of the ORCA BCG agrees with the published position of the maxBCG cluster to within 0.5 arcseconds.

2 *Cluster-galaxy match*: the position of an ORCA cluster member not classified as the BCG agrees with the maxBCG position to within 0.5 arcseconds.

3 *Cluster-centroid match*: We assess the statistical significance of locating an ORCA cluster centre between [θ, θ+dθ] from a maxBCG cluster position relative to a random distribution. We define a threshold θ_{max} where for $\theta < \theta_{max}$ the probability of finding an ORCA cluster centre exceeds the probability of finding a random point. The random data comprised 10 runs each with 1050 points. The frequency of random points around maxBCG cluster centres was determined in 10 radial bins of θ (with constant dθ) out to the maximum angular extent of the maxBCG clusters. From this analysis we adopt $\theta_{max} = 2.77'$; cluster-centroid matches are therefore separations of ORCA and maxBCG clusters less than this θ_{max}.

4 *Marginal matches*: Cluster centres with angular separations within the maximum angular extent of the maxBCG cluster, but with $\theta > \theta_{max}$ are classed as marginal matches

5 *Unassociated clusters*: ORCA clusters for which none of the above conditions are satisfied.

We find 8/22 BCG matches (1) to the maxBCG catalogue. There are no instances of a match in category (2), and 8/22 cluster centroid matches (3). A further 2/22 cluster centroids are considered marginal detections outside of θ_{max}=2.77′ (4). Of the 97 ORCA clusters, we find 78 are unassociated with any maxBCG entry. In total we therefore find an ORCA counterpart to 18 of the 22 clusters. The four maxBCG clusters without ORCA analogues are shown in Figure 3.19 with dashed circles, and are apparent as two pairs with small angular separation.

We note ORCA cluster *MGB J234341+00180.3* is situated between *BCG J234322+00190.6* and *BCG J234403+00130.6* (the eastern cluster pair with dashed red circles in Figure 3.19). Optical-band imaging (Figure 3.20) shows evidence of early type galaxies distributed in a filamentary chain, approximate comoving length $2h^{-1}\mathrm{Mpc}$, sampled by ORCA between the maxBCG detections.

The other pair (*BCG J234106+00120.4* and *BCG J234122+00190.0*) may be part of an elongated structure sampled by both the four maxBCG entries in that area and also by the ORCA detector. Figure 3.21 shows the ORCA cluster *MGB J234105+00180.3*. This cluster centre, situated between the two maxBCG clusters, matches the centroid of a NORAS cluster to within $0.4'$, with an uncertainty of $\sim 1'$ in the X-ray source.

Overall, we find very good agreement with the maxBCG catalogue of clusters, detecting 81% of their entries in the survey region, rising to 100% when taking into account how the different algorithms handle systems that by eye resemble filamentary structure.

3.5.2 X-ray detected clusters

We use cluster data from the *ROSAT* All Sky Survey-derived (RASS; Voges et al., 1999) NORAS (Böhringer et al., 2000) and BCS catalogues (for the latter, both main and extended catalogues; Ebeling et al., 1998, 2000), the XCS (Mehrtens et al., 2011) and BLOX (Dietrich et al., 2007) from *XMM-Newton*, and CHaMP (Barkhouse et al., 2006) from *Chandra*[10]. We combine these datasets to form a full-sky X-ray catalogue consisting of

[10]The naming prefixes adopted for these catalogues are: NORAS: *RXC*, XCS: *XMMXCS*, BLOX: *BLOX*, CHaMP: *CXOMP*.

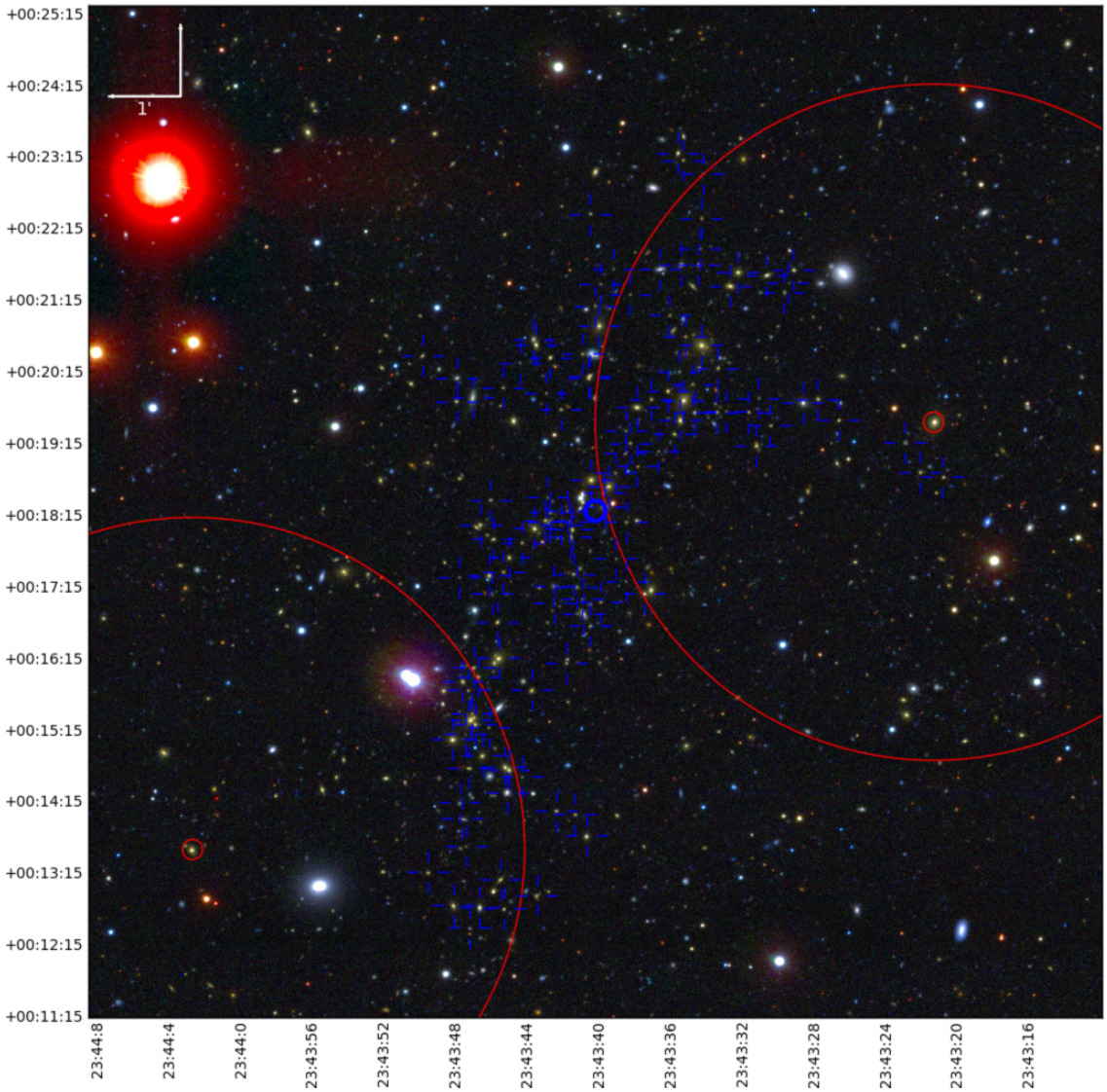

Figure 3.20: Stripe 82 cluster *MGB J234341+00180.3* is an extended system detected between two `maxBCG` clusters (*BCG J234322+00190.6* and *BCG J234403+00130.6*). For clarity, we have not plotted the Voronoi grid, but the cluster members are marked with blue cross-hairs. The `maxBCG` clusters are shown in red, with the central positions noted by the two smaller circles, and the larger circles corresponding to radii of $1h^{-1}\mathrm{Mpc}$ based on the photometrically-estimated cluster redshift from Koester et al. (2007a).

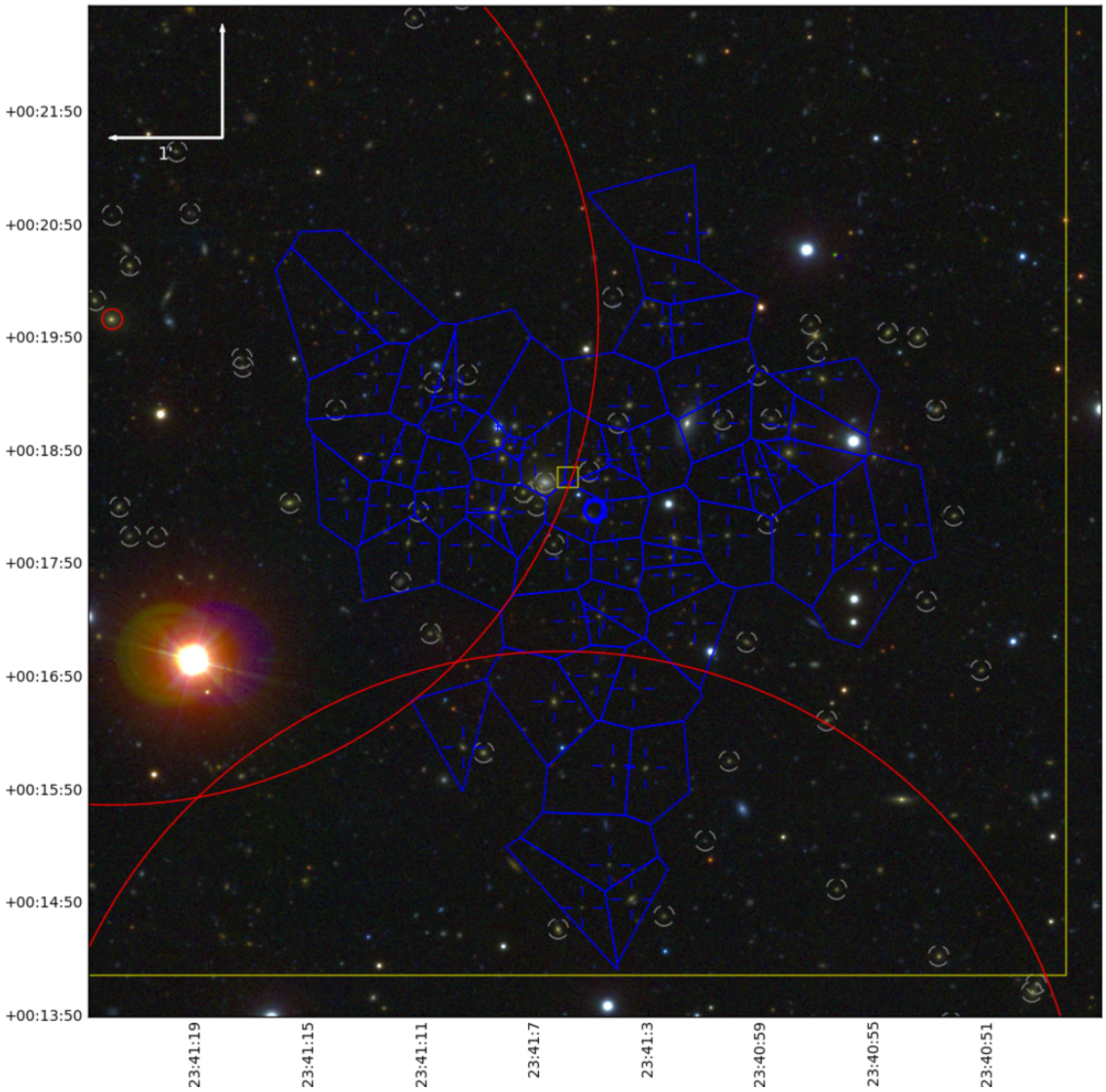

Figure 3.21: Stripe 82 cluster *MGB J234105+00180.3*: an ORCA detection between two maxBCG clusters and on top of an X-ray cluster position. Members and their Voronoi cells are marked in blue, the thick circle indicating the estimated cluster centre. Grey dashed circles are *associate cluster members* arising from multiple detections of this cluster (§3.3.6). Red data indicate the location of maxBCG clusters *BCG J234122+00190.0* and *BCG J234106+00120.4*, with larger circles indicating a $1h^{-1}\mathrm{Mpc}$ radius, smaller circles the BCG positions. Yellow data indicate the NORAS X-ray cluster *RXC J2341.1+0018*; the half-length of the large square corresponds to $1h^{-1}\mathrm{Mpc}$ based on the cluster redshift, the small square noting the X-ray position, uncertain to approximately 1′. The X-ray-ORCA centroid separation is approximately $0.4'$.

1463 unique clusters, taking care to identify any duplicate detections. From this catalogue, there are 58 X-ray clusters within the full 272.5 square degree area covered by Stripe 82, and two within the 7 square-degree sample studied here.

Approximately 80% of the 58 have redshift data, with sources measured out to $z = 1.19$ and a median redshift of $z = 0.27$. Blue squares in Figure 3.19 show the position of the two matches in the region we study here. The westernmost X-ray cluster, *RXC J2337.6+0016* (also detected in the flux-limited Brightest Cluster Sample, Ebeling et al., 1998) is the X-ray counterpart to *ACO2631* (Abell et al., 1989) and has a redshift of 0.2780 (Crawford et al., 1995). The X-ray position coincides with the `ORCA` detection of this system (*MGB J233740+00160.2*; z=0.2571) at a separation $(\Delta\theta, \Delta z)$ of $(0.1', 0.021)$. The easternmost X-ray cluster (*RXC J2341.1+0018*) with a redshift of z=0.2766 (Katgert et al., 1998, misidentified as *ACO2644*) was originally optically identified in Goto et al. (2002) and Lopes et al. (2004), and is in close proximity to *MGB J234105+00180.3* (z=0.2588), with $(\Delta\theta, \Delta z)$=$(0.4', 0.018)$. This latter match also appears to straddle two `maxBCG` clusters in the same region as the potentially elongated structure discussed in §3.5.1.

3.6 Mock survey cluster catalogue

In this section, we describe the application of `ORCA` to a mock PS-1 lightcone. Theoretical simulations allow one the luxury of comparing clusters detected by the algorithm (`ORCA` clusters) to the galaxy membership of dark matter haloes (hereafter ΛCDM clusters). Simulated galaxies are allocated to dark matter haloes using the Bower et al. (2006) semi analytic model. This approach makes the assumption a satellite galaxy is stripped of hot gas immediately following accretion onto a large halo. Star formation is halted after the cold gas reservoir is depleted, and the galaxy joins the red sequence. Coupled with AGN feedback, this prescription reproduces the observed bimodality in galaxy colours. However a known flaw, the rate of gas depletion, results in redder than observed satellite galaxies. Recent treatments of ram-pressure stripping (e.g., McCarthy et al., 2008) hope to improve understanding of the transition to early-type galaxies with improved semi-analytic models (Font et al., 2008; Benson and Bower, 2010).

Although mock surveys are inaccurate realisations of the universe (see Hilbert and White, 2010, for an example in a cluster detection context), they can nevertheless serve as

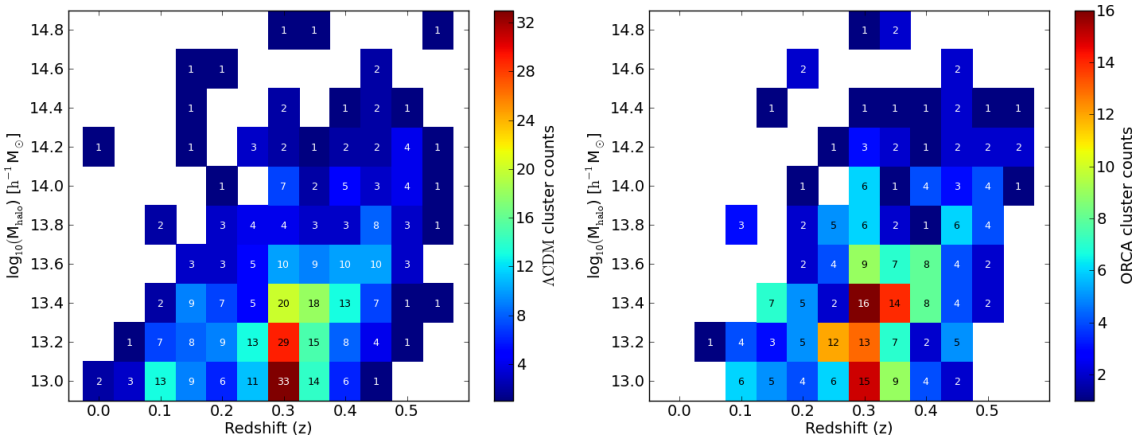

Figure 3.22: The number of ΛCDM (*left*) and ORCA (*right*) clusters detected in the mock survey as a function of redshift and halo mass.

self-consistent tests of the detector. We emphasise, however, there is little merit in comparing mock cluster detections with those in survey data until models can reproduce the observed group and cluster galaxy population with more fidelity. An at-a-glance impression of the difference in colour distribution can be seen in the colour-magnitude diagrams shown in Figures 3.10 and 3.12. Despite the same colour, magnitude and redshift limits, there is a clear difference in the colour distributions of the mock and real cluster. Because the colour data are more compressed in the mock CMRs, we opt to explore the same colour range as we did with the Stripe 82 survey.

3.6.1 Producing ΛCDM and mock ORCA cluster catalogues

The magnitude limits applied to the lightcone created a source catalogue of 80,536 mock galaxies. We use these galaxies to build our model ΛCDM cluster population. Because the algorithm relies on the detection of colour-magnitude ridgelines, we do not want to include ΛCDM clusters without detectable sequences[11]. We therefore construct the model cluster list from galaxies selected in the same photometric filters used by the detector. ΛCDM clusters may therefore also be detected multiple times. We group together any ΛCDM clusters with common halo identifiers, and select those with the highest *reduced fluxes* as the "best" model clusters. In each ΛCDM cluster, we calculate the approximate centre from cluster member positions. We find in some cases that haloes are

[11]The characterisation of galaxy clusters without red-sequences is, however, a very interesting study not covered by the work presented here.

assigned galaxies significantly separated from the majority of members in position or velocity space. Such galaxies are identified by rejecting 3σ deviations from a bootstrap-estimated median galaxy-centroid distance. Following outlier ejection, we find the resultant cluster sizes agree well with the virial radii of the host haloes. We set a minimum cluster mass limit by selecting ΛCDM clusters hosted by haloes with $M_H \geq 10^{13}h^{-1}M_{\odot}$. We find in total 414 ΛCDM clusters and in the *left* panel of Figure 3.22, we plot the number of model clusters as a function of redshift and halo mass.

Except for the revised parameters listed in Table 3.3, the detector ran as described in §3.3. We found a total of 305 ORCA clusters with $M_H \geq 10^{13}h^{-1}M_{\odot}$; at $M_H \geq 10^{14}h^{-1}M_{\odot}$ the counts are more equal. As with the ΛCDM clusters, the clusters are most commonly identified at $\mathrm{z} \sim 0.3$ (see the *right* panel of Figure 3.22). However, the tests we describe in the following section will highlight how well ORCA performs over the entire redshift range.

Figure 3.23 shows a simple comparison of the two catalogues by plotting both sets of clusters residing in haloes $M_H \geq 10^{13.5}h^{-1}M_{\odot}$ out to $\mathrm{z} = 0.6$ (the highest cluster redshift in the SDSS cluster catalogue). Grey circle centres denote the position, and their radii the maximum member-cluster centre distance of ΛCDM clusters. Blue and red cells represent ORCA clusters detected in {*g-r*,*r-i*} and {*r-i*,*i-z*} respectively.

3.7 Estimating the cluster selection function

To determine how well the detector recovers and characterises the mock clusters, we illustrate here three simple tests to quantify the detection performance.

3.7.1 Completeness

We define completeness as the number of detected haloes as a function of halo mass and redshift. A halo is detected if at least N_{min} galaxies are identified, even if they are shared between multiple ORCA clusters (for example, fragmenting a halo when the algorithm attempts to identify substructure). We compare this number to ΛCDM cluster counts (by definition unfragmented), with at least N_{min} members. The fraction of detected ΛCDM clusters can be seen in Figure 3.24, where we produce a grid of cells with sampling intervals of 0.05 in redshift and 0.2 in $\log_{10}$ halo mass. Because in some cases only a

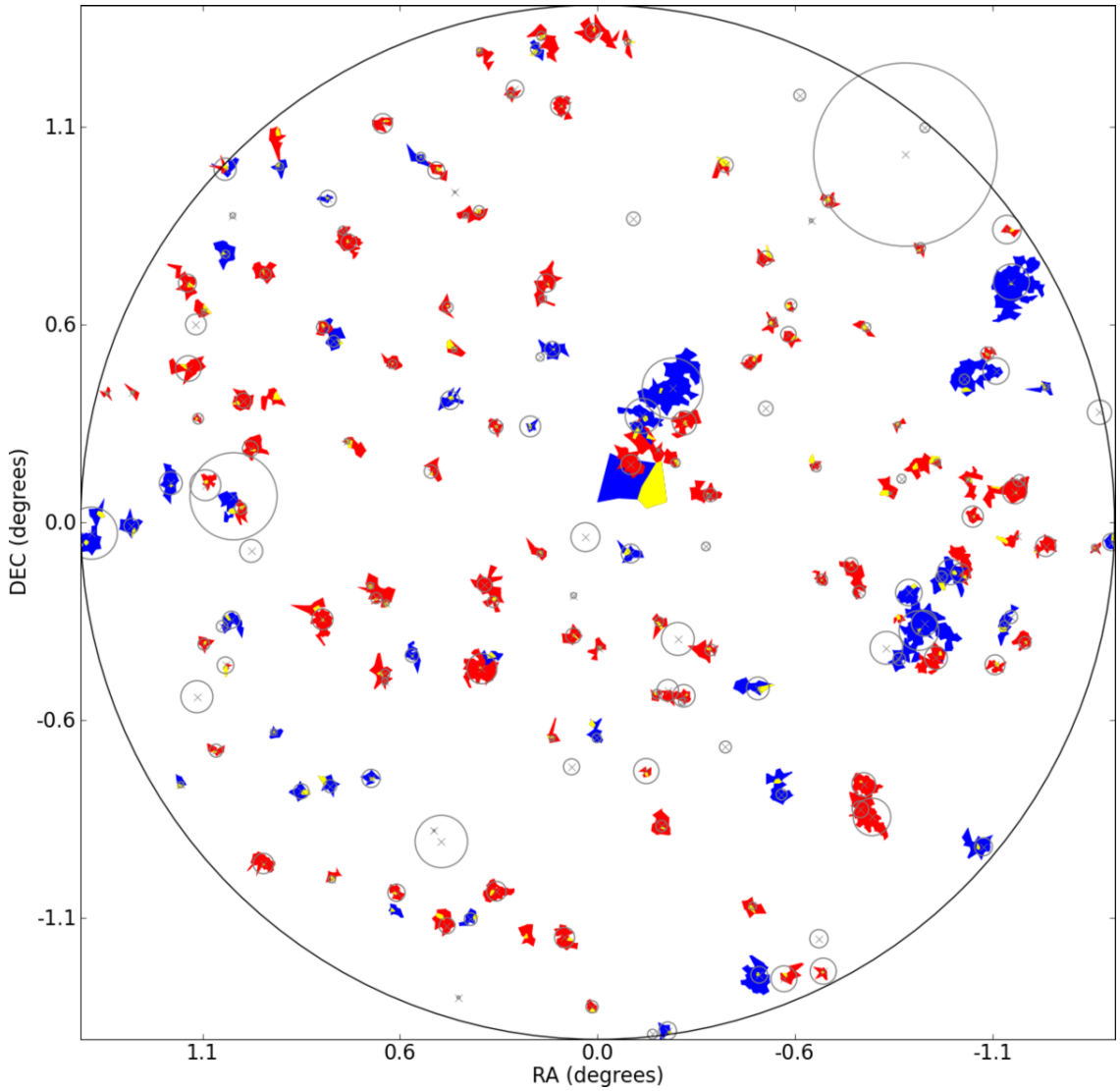

Figure 3.23: Clusters in haloes of mass $\geq 10^{13.5}h^{-1}M_{\odot}$ from the mock ORCA cluster catalogue (cells) and the ΛCDM catalogue (circles). Cell colours correspond to clusters detected in different colour pairs. Blue cells are clusters detected in the $\{g\text{-}r, r\text{-}i\}$ filter pairs, red are clusters detected in $\{r\text{-}i, i\text{-}z\}$. Yellow cells indicate the BCG of each cluster. Crosses denote the ΛCDM cluster centre, and circle radii indicating the angular distance between the centre and most distant member.

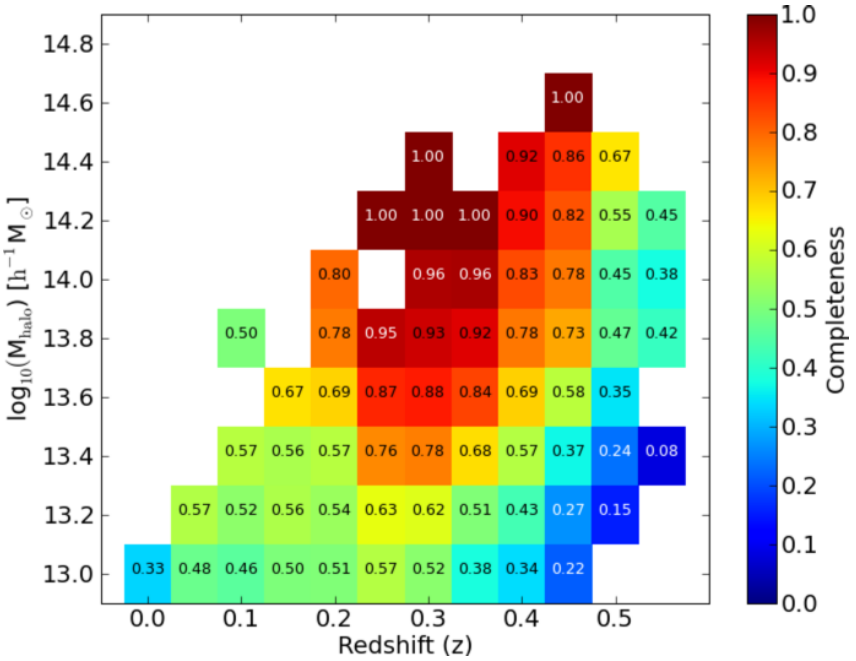

Figure 3.24: Completeness of mock ΛCDM clusters. The fraction of correctly detected clusters from the ORCA catalogue as a function of halo mass and redshift. The white regions indicate where there were no ΛCDM clusters in that bin.

few detections occupy each cell, some regions will suffer from shot noise. We smooth the data using a 3×3 grid so the completeness for a given cell is the mean completeness over this region. Empty regions in Figure 3.24 therefore indicate where either no ΛCDM clusters exist or too few clusters are found to reliably calculate the completeness (we set a threshold of at least five clusters detected over the 3×3 grid). Between $0.1 \leq z \leq 0.4$, the detector attains at least 68% completeness for halo masses above $10^{13.6} h^{-1} M_{\odot}$, and is over 90% complete in halo masses exceeding $10^{14.3} h^{-1} M_{\odot}$. This compares favourably with the maxBCG algorithm applied to mock simulations, where Koester et al. (2007b) report $> 90\%$ completeness between $0.1 \leq z \leq 0.3$ for $M_H \geq 10^{14.3} h^{-1} M_{\odot}$ with clusters containing at least 10 members (cf. $N_{min} = 5$ in this study). Applying the completeness definition and the same selection criteria as that study, the ORCA detector is $> 90\%$ complete down to a halo mass of $10^{13.8} h^{-1} M_{\odot}$. These results also compare well to the Voronoi Tessellation completeness of the 2TecX (van Breukelen and Clewley, 2009) algorithm, either matching or exceeding their stated completeness for $M_H = 10^{13.7}$ and $10^{14} h^{-1} M_{\odot}$ up to our redshift limit. The C4 (Miller et al., 2005) algorithm analysed

SDSS DR2, an earlier release of the Sloan catalogue than that searched with `maxBCG`. `C4` was tested on mock data, and revealed a completeness of approximately 75% between $0.093 \leq \mathrm{z} \leq 0.107$ for halo masses up to $10^{13.8}h^{-1}M_{\odot}$, compared to an `ORCA` completeness of 50% down to the same mass. However, due to the stringent detection conditions used by the `C4` algorithm, clusters are only detected to $\mathrm{z} \leq 0.17$; in the deepest redshift bin analysed ($0.118 \leq \mathrm{z} \leq 0.128$), the completeness drops to approximately 20% at the same mass. Our closest completeness estimates here vary between 67% and 78%.

At higher redshifts there is a decline in completeness where there are only a few members brighter than the magnitude limit, reducing the algorithm sensitivity to distant clusters. This effect is more apparent among the lower mass haloes. At high redshift ($\mathrm{z} > 0.4$) and low mass ($M_H \leq 10^{13.3}h^{-1}M_{\odot}$) there are 12 ΛCDM clusters, but the detector identifies only two of these. We also note a local incompleteness at $\mathrm{z} \leq 0.08$. Arising from our choice of probability threshold ($\mathrm{P_{thresh}}$), too few overdense cells are selected in filters featuring low signal-to-noise clusters. The photometric filters best suited to detecting local, relatively blue clusters have galaxy populations dominated by the blue cloud component of the colour-magnitude relation. Successful detections in this crowded field are compounded by the larger scale-size of more local clusters such as the local ($\mathrm{z} = 0.03$) seven-member group at the north-western boundary of the catalogue in Figure 3.23. Under these circumstances, it becomes unlikely cluster Voronoi cells share common vertices, restricting potential membership links between them.

We classify spurious detections in the mock cluster catalogue as those clusters where each member belongs to a different halo. Of the 305 `ORCA` clusters, only two fit this description, suggesting a spurious detection rate (0.7%) consistent with tests performed in §3.4.6.

3.7.2 Stellar mass accuracy

Stellar mass accuracy is the stellar mass of an `ORCA` cluster relative to that of the ΛCDM cluster belonging to the same halo. Because the algorithm may split the halo galaxies into multiple clusters, we combine the mass of all `ORCA` clusters sharing the same halo. In ΛCDM clusters with up to ∼12 members (approximately 75% of the catalogue), over half of the total cluster stellar mass comes from the two most massive galaxies. The efficient

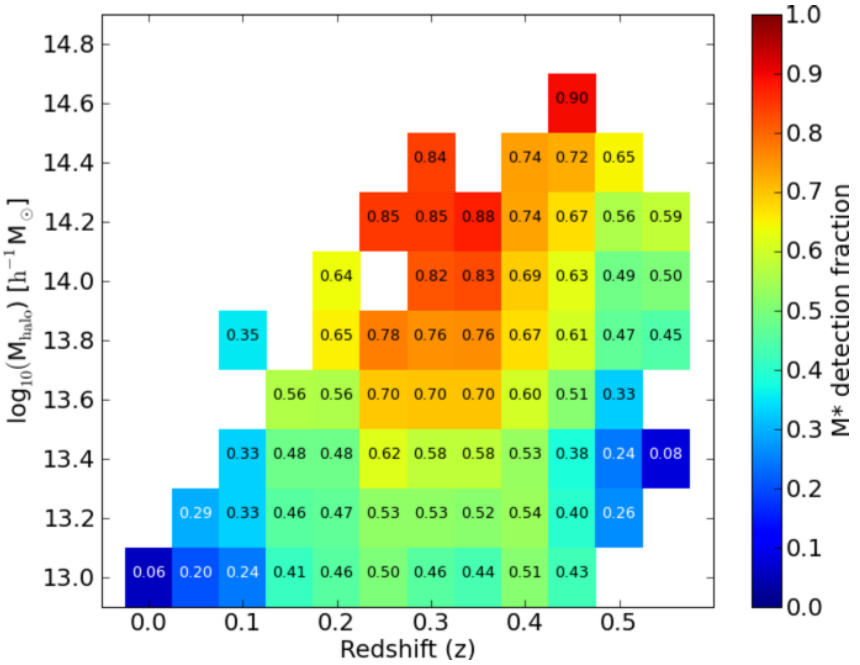

Figure 3.25: Stellar mass accuracy. The fraction of recovered stellar mass in mock clusters as a function of halo mass and redshift.

detection of these galaxies is therefore essential in gaining accurate estimates of cluster stellar masses. The stellar mass accuracy for each ΛCDM cluster is $A_* = M_{cl}/M_{true}$, where M_{cl} is the stellar mass of all ORCA cluster members registered to the ΛCDM cluster's halo. We apply the same gridding technique discussed in the previous section, requiring at least 5 clusters in a grid to define a reliable A_*. As Figure 3.25 shows, between $0.1 \leq \mathrm{z} \leq 0.4$ the algorithm recovers over half of the cluster stellar mass for systems with halo masses of at least $10^{13.4} h^{-1} M_\odot$. This recovery fraction improves with increasing mass, reaching 90% in some cases. Both local and distant clusters suffer from lower stellar mass estimates. For the former, higher levels of halo fragmentation (one halo being assigned to many ORCA clusters) result in galaxies lost to nearby systems with densities or memberships too low to qualify as clusters. Those systems with redshifts $\mathrm{z} > 0.5$ tend to be unfragmented but contain fewer members, causing an underestimation of cluster stellar mass. The stellar mass accuracy at the median redshift of the survey ($\mathrm{z} = 0.33$) remains above 50% down to halo masses of $10^{13.2} h^{-1} M_\odot$, and above 75% from masses of $10^{13.8} h^{-1} M_\odot$, suggesting the detector performs well in estimating the true cluster stellar mass content.

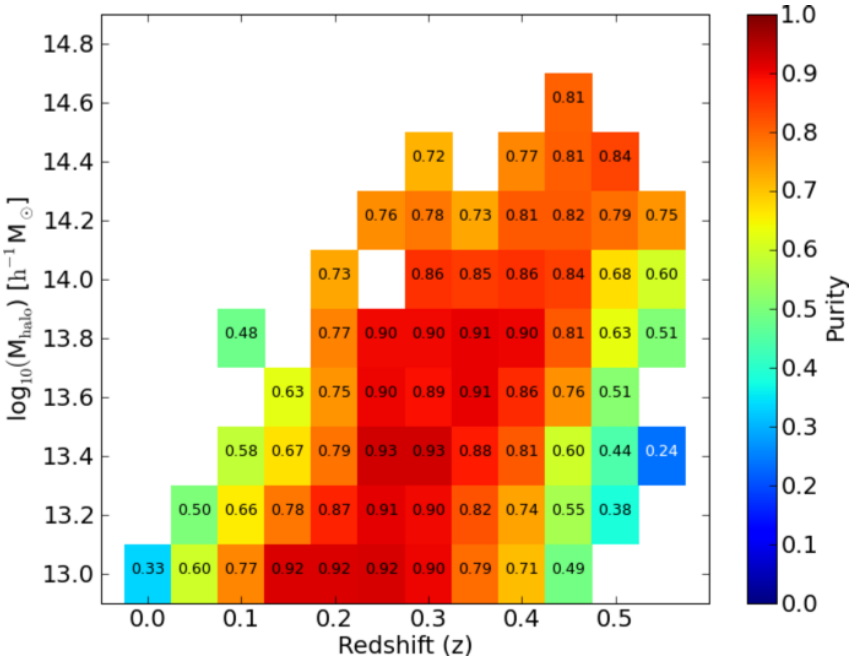

Figure 3.26: The purity of ΛCDM clusters detected by the ORCA algorithm. Low values indicate where clusters have included a large number of contaminating galaxies not belonging to the halo.

3.7.3 Purity

As discussed in §3.7.1, a halo is detected by the algorithm if it finds at least N_{min} members that have been allocated to ORCA clusters. For a cluster with 7 members, the distinction between a cluster containing 5 halo galaxies and 2 interlopers and one containing 7 halo galaxies provides a measure of cluster purity. We define purity as the fraction of galaxies ORCA assigned to the cluster that are members additionally belonging to the host halo. This description differs from that defined in the evaluation of the C4 algorithm (Miller et al., 2005), but is in line with the purity described by Koester et al. (2007b). However, in our case we decide not to adopt a threshold above which a cluster is considered pure, instead directly assigning each cluster a purity fraction. Figure 3.26 shows the purity of ORCA clusters with varying redshift and halo mass, the gridding method here being the same scheme introduced in §3.7.1. ORCA clusters are at least 70% pure at the median redshift of the survey over all halo masses. The purity appears to drop at higher redshifts, attributed to faint but genuine cluster members being replaced by brighter contaminants that lie on the cluster sequence. Relative to the completeness and

stellar mass estimates, cluster purity is not as sensitive to halo mass. This is most likely a consequence of the membership incompleteness discussed in §9. Because peripheral members are less likely to be in Voronoi cells tagged as statistically significant, the inclusion of interlopers at cluster edges is reduced. As in the previous section, increased halo fragmentation drives the local drop in purity, serving to increase the contamination fraction by distributing the halo galaxies among local clusters and systems failing to achieve cluster status.

3.8 Summary

We present and demonstrate a new cluster detection algorithm based on red-sequence cluster searches, the detection of overdensities using Voronoi Tessellations, and connection of galaxies into clusters with a Friends-of-Friends algorithm. With this approach, we make only two assumptions about the systems we are looking for: that they have detectable red-sequences, and are overdensities in the projected plane of the sky.

We calibrate the photometric selection filters to a rich Abell cluster found in SDSS data, and find that recovery of members from both this large cluster and a small group is largely insensitive to the choice of two algorithm parameters controlling the behaviour of the algorithm. When applying the algorithm to a sample of SDSS Stripe 82 galaxies with four bands, we find 97 clusters. Based on spectroscopic and photometric redshifts, we estimate these clusters are detected out to $\mathrm{z} = 0.6$, and the catalogue has a median redshift of $\mathrm{z} = 0.31$. We perform false-positive tests suggesting the spurious detection frequency is below 1%. Tests on the catalogue suggest the detector is robust to sparsely sampled cluster fields and is not overly sensitive to survey edges. In comparing our data to existing optical and X-ray clusters, we find good agreement with the `maxBCG` and RASS catalogues in the same region.

We go on to test the performance of the detector with a mock survey generated from a semi-analytic galaxy formation model. In comparing the `ORCA` cluster detections to those generated from halo membership data, we make a quantitative assessment of the detector performance. The algorithm identifies 305 clusters, whilst the simulation produces 414 down to a halo mass of $10^{13}h^{-1}M_{\odot}$. At the median redshift of the catalogues (both $\mathrm{z} = 0.33$) we find `ORCA` is 75% complete down to a cluster halo mass of $10^{13.4}h^{-1}M_{\odot}$ and

is able to recover approximately 75% of the total stellar mass for clusters in haloes of at least $10^{13.8}h^{-1}M_{\odot}$.

We have demonstrated this algorithm is capable of identifying clusters in both real and simulated data with minimal assumptions as to the nature of clusters. In combining comprehensive colour scans to search for cluster red-sequences with Voronoi diagrams to estimate surface densities, we avoid making model-dependent decisions about what a cluster is. Cluster redshifts arise as a consequence, not condition, of our detection, affording additional freedom from model SEDs and the uncertainties inherent in photometric redshift data spanning the depths, fluxes and areas set to be commonplace in next-generation galaxy catalogues. This detector can be used in any survey where there are at least two photometric bands, but is most powerful when applied to multi-colour surveys such as the forthcoming Pan-STARRS surveys.

Chapter 4

The ORCA Stripe 82 cluster catalogue

4.1 Overview

In chapter 3, we introduced and demonstrated the ORCA cluster detection algorithm. By testing our prescription using both real and mock data, we were able to verify first that ORCA is able to detect clusters in photometric data, and subsequently that we understand the performance of the detector. Whilst this testing process is vital to producing homogeneous, well-defined cluster catalogues, we must also develop the algorithm to allow searches through a variety of (often high data-volume) photometric surveys. In particular, it is important that detection algorithms such as ORCA can scale to the areas covered by the Pan-STARRS 3π survey.

In this chapter, we first describe the changes made to ORCA in preparation for the algorithm's application to a "large-area" survey. Both the volume of data and large parameter space explored by ORCA requires careful consideration of the memory and processing constraints. Following these developments, we present a catalogue of 4098 photometrically selected galaxy clusters with a median redshift $z_{\rm med}$=0.32 in the 270 square-degree 'Stripe 82' region of the SDSS. This survey covers the celestial equator in the Southern Galactic Cap ($-50^\circ < \alpha < 59^\circ$, $|\delta| \leq 1.25^\circ$). Owing to the multi-epoch SDSS coverage of this region, the *ugriz* photometry is $\sim$2 magnitudes deeper than single scans within the main SDSS footprint. We exploit this to detect clusters of galaxies using ORCA. 32% of the clusters have at least one member with a spectroscopic redshift from existing public data (SDSS Data Release 7, 2SLAQ & WiggleZ), and the remainder have a robust photometric redshift (accurate to $\sim$5–9% at the median redshift of the sample). The weighted median of member redshifts provides a reasonably accurate estimate of

the cluster redshift. The cluster catalogue is publicly available for exploitation by the community to pursue a range of science objectives.

In §4.2 we describe the steps taken to improve the running of the ORCA detector. In section §4.3 we describe the construction of the data catalogue and how it was subdivided; in §4.4 we apply the detection algorithm to this data and discuss the contents of the catalogue. In §4.5 we investigate the blue limit established in chapter 3 and extend the search for clusters bluewards. We compare our measurements of the red sequence slope and BCG colour tracks to model predictions in §4.6 and §4.7, and summarise our findings in §4.8.

4.2 Preparing for wide-area cluster searches

Because our cluster detection philosophy only relies on two core assumptions and makes few constraints on the input data, ORCA is expected to explore simultaneously through large data volumes and a wide, multi-colour parameter space. During this search a series of processor-intensive calculations are made to identify putative clusters and their members. A considerable challenge to this approach has been to remain computationally feasible in the face of large data volumes. Because spatial, colour and flux data are required for each galaxy, careful auditing of the memory footprint was required throughout development of the algorithm, in addition to balancing this usage with slower disk access. Within ORCA, many of the iterative operations will cause progressively longer run times as the number of galaxies increases. We identify the two most significant performance bottlenecks in the ORCA detection process as:

- Sequential processing of single-colour photometric filters (e.g. $\{g\text{-}r\}$) followed up with dual-colour (e.g. $\{g\text{-}r,r\text{-}i\}$) analysis (see §3.3.3)

- The percolation algorithm connecting neighbouring Voronoi cells in order to build cluster membership.

To address these particular aspects of the detection process, two important changes to the algorithm were made.

4.2.1 Parallelisation

Various parts of ORCA were parallelised. In particular, where the same operations were performed on a series of photometric filters, rather than have the algorithm run these serially the multi-processor (via symmetric multiprocessing; SMP) capabilities of the host machine are exploited. We achieve this using a self-designed wrapper to the Parallel Python[1] (PP) project: a job server distributing tasks to either CPUs or across a cluster of computers. ORCA has been predominantly used in a single-system, multi-core environment, but has also been tested on a computer cluster. Advantages of SMP over computer cluster-based deployment are two-fold: i) The volume of data transmitted between server and client, although variable, can become bandwidth-limited over ethernet-based clusters ii) For a collection of parallel jobs reading the same data from disk, parallelisation across a cluster of computers does not benefit as much from disk caching (separate nodes must each read the data in full). With future developments aimed at reducing disk access and the data volumes transmitted for each job, ORCA deployment over computer clusters may be improved. Examples of procedures implementing PP are:

- The application of different photometric filters to an identical source catalogue.
- Generating a series of independent Voronoi tessellations.
- Applying the friends-of-friends percolation routine to identify clusters from this series.
- The calculation of galaxy cluster richnesses.

4.2.2 Improved percolation code

One drawback in adopting Python as the ORCA development language is the execution speed for tasks requiring nested-loop or processor-intensive operations - these tend to be slower than equivalents in pre-compiled languages such as C. This is particularly evident in the ORCA percolation procedures, which reduce to a computational problem akin to the calculation of the two-point correlation function. Using a "brute-force" implementation: for every point in a set of N, $N-1$ criteria are checked prior to analysing the

[1] http://www.parallelpython.com

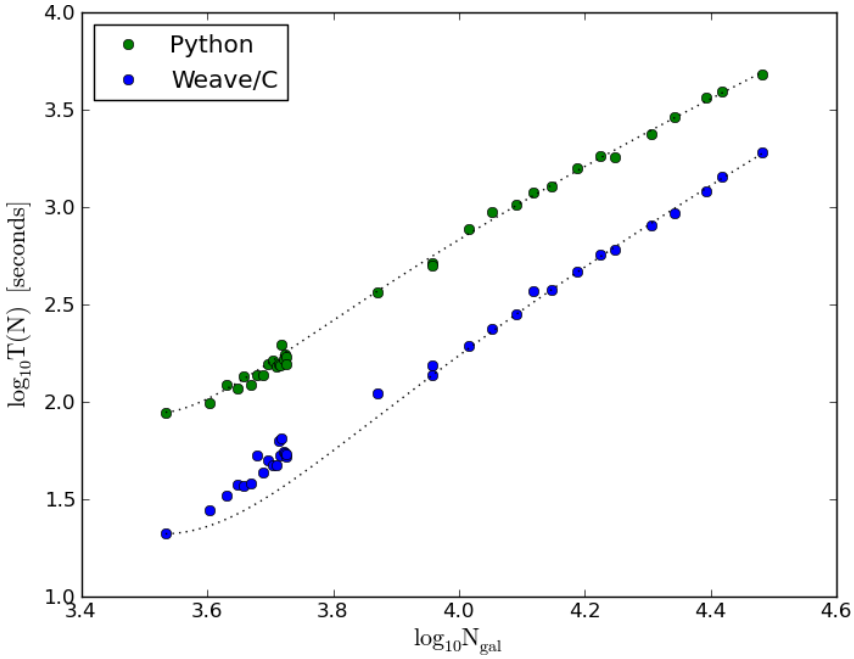

Figure 4.1: Comparative performance of two percolation codes. The blue line corresponds to the time taken to perform the `ORCA` friends-of-friends percolation procedure using the original (`Python`) code. The green line (`C`) shows results from improved code that uses the `Python Weave` module to embed C-code into `ORCA` .

next point (in `ORCA`'s case, Voronoi cell vertices). These approaches tend to run in order $\mathcal{O}(N^2)$. Our solution improves the speed of frequently used, nested-loop `Python` functions by converting them to `C` and pre-compiling them. A new percolation function was written in `C` and integrated into `ORCA` with the `Python Weave` module[2]. `Weave's` "inline" feature allows the insertion of `C` code directly into the `Python` function. Compiled only once, future calls to this code fragment are executed much faster than the original code.

Figure 4.1 shows the performance gain from adopting this new approach. Across sample sizes representative of the galaxy numbers `ORCA` analyses[3], we measured the time taken to complete the friends-of-friends percolation under the original `Python` implementation (*green*) and the improved `Weave` code (*green*) when applied to the mock

[2] `http://www.scipy.org/Weave`

[3] These numbers correspond to the peak number of galaxies selected across all photometric filters rather than the total number of galaxies in the survey.

lightcone introduced in §3.2. These tests were performed on an 8-core AMD machine with a 2.5GHz clockspeed. Although this remains a "brute-force" approach (alternatives include kd- or binary-trees), the improved performance does not justify additional optimisation.

To assess how completion time scaled with N, we fit a model of the form $\mathrm{T(N)} = \mathrm{aN^b}$ to the data. These fits can be seen as dotted lines in the Figure. The `Python` approach evolves as $\mathcal{O}(\mathrm{N}^{1.49})$, whilst the `Weave` algorithm evolves as $\mathcal{O}(\mathrm{N}^{1.77})$. However, the ratio $\mathrm{a_{Python}/a_{Weave}} \simeq 5$, so the latter approach is quicker (by 0.5 orders of magnitude in the range we explore), with the slower algorithm only becoming competitive for completion times T(N)$\sim$6.5 days. These results suggest the need for some sensible limit to the number of galaxies analysed in one step. Because it is computationally inefficient to process extremely large datasets (notwithstanding memory considerations), we must therefore spatially divide input data into smaller subsets. Despite the computational overhead in stitching together these partitions, this strategy will permit the analysis of data from projects such as the 3π Pan-STARRS survey. We discuss the partitioning techniques in both §4.3.2 and §5.6.2.

4.3 The Stripe 82 data catalogue

In lieu of Pan-STARRS data (currently being accumulated; survey mode having commenced in May 2010), we have put `ORCA` to immediate use on other photometric surveys. Here we present our detection efforts on Stripe 82 - a sub-region within the SDSS which was re-visited many times. This deeper equatorial strip, $|\delta| \leq 1.25^\circ$, spanning $-50^\circ < \alpha < 59^\circ$ in right ascension, totals approximately 270 square degrees. The co-addition of 47 and 55 strip scans (corresponding to the southern and northern parts of the Stripe) results in a catalogue of objects $\sim$2 magnitudes deeper than an individual scan.

4.3.1 Data selection

As in chapter 3, we used the SDSS Catalogue Archive Server (CAS)[4] to extract *griz* model magnitude photometry from the `PhotoObj` table for all galaxies in the Stripe 82 co-add.

[4] `http://casjobs.sdss.org`

We enforce a magnitude limiting range: $14 < r \leq 24$ and to eliminate stellar contamination we stipulate an additional minimum offset between the point-spread function fit magnitude and model magnitude: $(r_{\rm PSF} - r_{\rm model}) > 0.05$. All photometry is corrected for Galactic extinction using the relevant '`extinction`' table value (Schlegel et al., 1998). To remove overly de-blended or saturated sources, and those near frame edges, we make use of the CAS `fPhotoFlags` parameter. Full details of the query can be found in Appendix B.

There are a total of 11,154,087 galaxies in the catalogue. For every galaxy, we determine whether a spectroscopic redshift is available from either the SDSS Data Release 7 (DR7), 2dF-SDSS LRG and QSO (2SLAQ; Croom et al., 2009)) or WiggleZ Data Release 1 (Drinkwater et al., 2010). To supplement these data, we also use two photometric redshift catalogues. These are discussed in more detail in §4.4.1.

4.3.2 Data partitioning

As recommended in §4.2.1, to improve computational performance the catalogue was partitioned into 44 smaller sectors, divided in right ascension. Each sector, of width $\sim 0.2^{\rm h}$, overlapped the adjacent sector by $3^{\rm m}$ to ensure no clusters were accidentally sliced in half and subsequently not detected in either of the two sectors. The algorithm was therefore required to account for the overlap zone between each sector. Within this zone, not only may clusters be detected more than once, but detections at the very edge of the sector may be clipped. We remedy this by devising a set of rules to describe when an overlap-zone cluster should be accepted, rejected or treated as a duplicate detection. Figure 4.2 shows two neighbouring Stripe 82 sectors and demonstrates how we split the overlap region (central in this plot) equally into three regions. The status of an overlap-cluster depends on the sector it was detected in and the zone it resides in within that sector. In this Figure, the circles denote cluster positions and projected sizes[5]. Circle colours correspond to the sector the cluster was identified in: *red* for left, *green* for right.

The *bottom* panel of Figure 4.2 shows the jurisdictions of the two sectors under our rule-set. Clusters outside the overlap zone are detected only one sector (the darker shades of *red* and *green*), and are added to the final cluster catalogue. All clusters within

[5]In this particular example, the merging of multiple detections has yet to be performed.

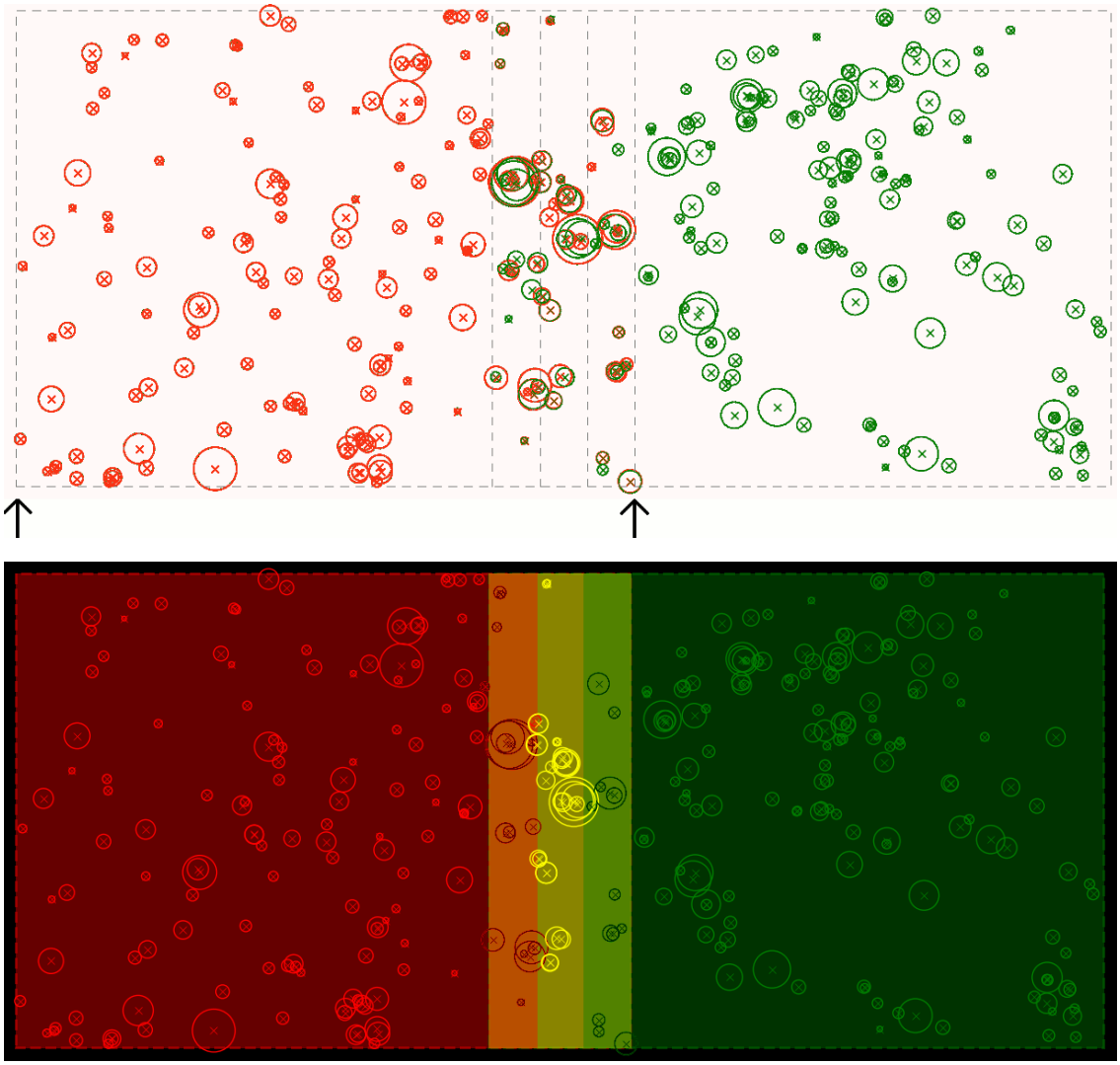

Figure 4.2: An example of the sector overlap scheme in the sub-divided Stripe 82 catalogue. Clusters are denoted by circles, those from the left(right) sector coloured red(green). In the *top* panel, we show all clusters detected in these two sectors The two arrows indicate the limits in RA of the left-hand sector, and the central vertical dashed lines indicate the three regions that comprise the overlap zone. The *bottom* panel shows the overlap scheme applied: the central yellow region shows clusters to be passed to the merging algorithm; clusters detected in the left-hand sector but located in the light-green zone have been removed. Equivalently, clusters detected in the right-hand sector by located in the light-red region have also been removed.

the central partition of the overlap zone (*yellow*) are designated potential multiple detections and are passed to the algorithm controlling the merge process (§3.3.6). Clusters in this zone with no nearby neighbours are treated as uniquely detected clusters and added to the final cluster catalogue. Those clusters located in the overlap region most distant from their sector (e.g. *"green"* clusters in the *light-red* region) are removed; genuine clusters should already have been identified in the nearer sector. This is demonstrated in the difference between the *top* and *bottom* panels, where in the latter we have removed these clusters.

After the detection of clusters in all 44 sectors, a merging routine stitches together all identified clusters, applying the above prescription to process overlapping areas, thus generating the final cluster catalogue.

4.4 The Stripe 82 ORCA cluster catalogue

From the Stripe 82 galaxy data, we detect a total of 4098 unique clusters with $\mathrm{N_{gal}} \geq 5$ members, assigning 59818 galaxies to clusters. We plot the angular and redshift positions of these systems in Figure 4.3. It is important to note other optical cluster-finding efforts in the Stripe 82 region.

Although ORCA is able to identify the majority of Abell clusters within the stripe (35/37), both the shallow depth and known projection issues (e.g. Lucey, 1983) within the Abell et al. (1989) catalogue suggest Abell completeness is not the best metric for detection performance. We provided a detailed comparison of ORCA and maxBCG (Koester et al., 2007a) cluster detections of 7deg^2 in §3.5. When we analyse the full Stripe 82 footprint, there are a total of 492 maxBCG clusters, of which >90% share an ORCA counterpart. Direct comparison of the relative efficacy of the two algorithms is unfair, since the maxBCG catalogue was applied to the shallower SDSS photometry prior to the multi-epoch Stripe 82 co-added data. As a result, we are able to detect more clusters (including groups of fainter systems), out to higher redshifts than in Koester et al. (2007a). Massive clusters, generally very prominent in projection, are likely to be easily detected even by very different techniques. Variations in the low-mass cluster yield may instead be a good discriminator between the methods, with the caveat that definitions of a galaxy cluster will vary. Restricting the ORCA catalogue to $0.1 \leq \mathrm{z} \leq 0.3$ to better match the maxBCG

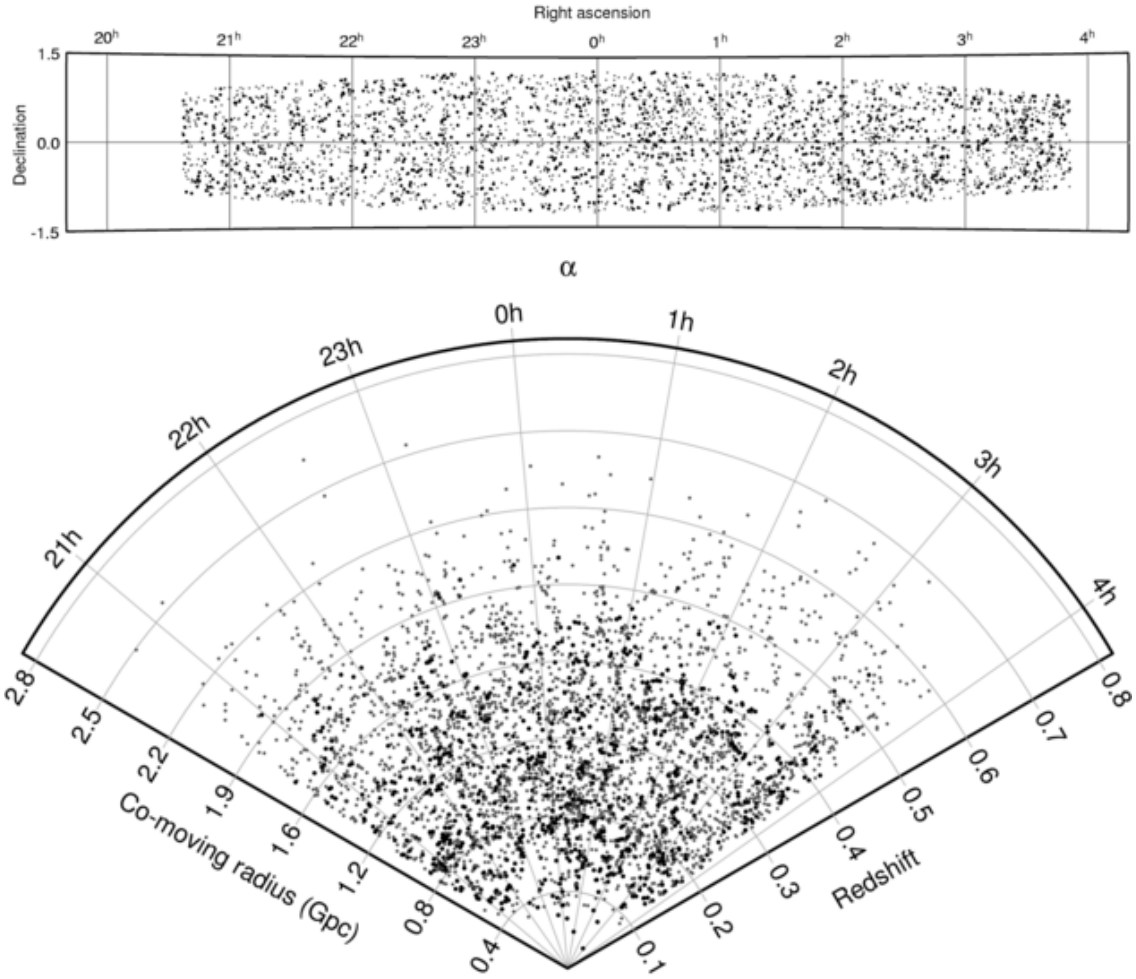

Figure 4.3: (adapted from Geach et al., 2011). Sky and redshift distribution of clusters detected in Stripe 82. The top panel shows the angular distribution (Aitoff projection) and the lower cone plot indicates the redshift distribution over the full right ascension range of the Stripe, projected in $\sim$2.5 degrees of declination. Larger points indicate BCGs with spectroscopic measurements (we describe redshift estimation of clusters in §4.4.1). We comfortably detect clusters out to $\mathrm{z} \sim 0.6$, with a handful of systems potentially detected at higher redshifts. Beyond this, the 4000Å break moves into the near-IR, and selection using the Sloan optical bands alone becomes inefficient.

redshift range, we detect 1794 clusters and groups with $\geq$5 members. Setting the minimum membership to $\geq$10 members however, we find 504 systems - comparable to the maxBCG surface density. This demonstrates (assuming membership scales with mass) that the low-mass regime is most important when comparing cluster catalogues.

To estimate the yield of maxBCG clusters from deeper Stripe 82 data, we turn to the Millennium simulation. Koester et al. (2007b) claim 90% completeness at a halo mass of $2 \times 10^{14} h^{-1} M_{\odot}$, whilst in §3.7.1 we calculate 90% ORCA completeness at a similar mass ($1 \times 10^{14} h^{-1} M_{\odot}$). In the z=0 Millennium Snapshot there are 1938(619) dark matter haloes with $\mathrm{M}_{200} \geq 1 \times 10^{14} (2 \times 10^{14}) h^{-1} \mathrm{M}_{\odot}$. One would therefore expect (1938/619)$\times$492=1540 clusters if maxBCG has the same sensitivity as ORCA . In the same redshift limits, ORCA identifies 1793 clusters, meaning we still observe $\sim$16% more clusters in the ORCA catalogue (assuming all Stripe 82 clusters at $\mathrm{z} < 0.3$ are detectable in the single-epoch SDSS data). Returning to the combined catalogue of X-ray cluster detections highlighted in §3.5.2, we noted 58 X-ray clusters were identified in Stripe 82. A preliminary cross-matching of these X-ray clusters to positions of ORCA systems determined that 24 of the 58 did not have optical counterparts in our catalogue. Investigating these further, 8 did not have redshift data, 6 have cluster redshifts beyond the z <0.6 range probed by the ORCA catalogue and 9 were, in principle detectable.

Of the latter group, we visually compare the two datasets. We find 3/9 were in reality detected by ORCA (in each case the interpolated cluster centre was just outside of the X-ray position). One particular X-ray system (*XMMXCS J021518.8-003813.2*) was undetected, but was close to a bright star that may have impaired the selection of cluster galaxies. This X-ray cluster has an estimated redshift (z=0.28) discordant with three nearby ORCA clusters at $0.39 < \mathrm{z} < 0.50$, and so is likely to be an entirely new system undetected by the optical algorithm. ORCA misses two X-ray clusters entirely: *RXC J0114.9+0024* (*ACO168* nearby at z=0.04, where ORCA is known to be incomplete) and *XMMXCS J030317.4+001238.4* at z=0.59 (also recorded amongst the 8 without redshift data as *BLOX J0303.3+0012.8*). The remainder comprise a collection of X-ray detected systems within a heavily targeted field at (α,δ)$\simeq$(46.64°, -0.08°) that was sampled only sparsely by ORCA .

From the 8 without measured redshifts (the majority of which were BLOX Dietrich et al., 2007, clusters), on closer inspection 3 were found to have ORCA counterparts. There

were two instances where it appeared the X-ray catalogue had either detected two components of a merging system or two substructures within a cluster (one such example is the *CXOMP J010610.3+005126-CXOMP J010607.0+004943* pair). There were 3 systems from a heavily sampled field (10 measurements between BLOX and XCS) that ORCA did not find counterparts for, and 1 cluster (*XMMXCS J021020.7-000706.5*, a small group of galaxies close to a bright star) that ORCA missed.

The serendipitous nature of some surveys comprising the X-ray catalogue mean the combined dataset will have a position-dependent depth that is hard to fairly compare to the ORCA data. There are ORCA optical counterparts to all 8 X-ray clusters in the wide-area NORAS (Böhringer et al., 2000) catalogue, and similarly matches to all flux-limited BCS (Ebeling et al., 1998) clusters except for the aforementioned *ACO168* system.

In the following sub-sections we describe the main properties of the Stripe 82 cluster catalogue including the cluster-derived observables such as cluster redshift estimates and optical richness.

4.4.1 Redshift estimation

Red-sequence galaxies lend themselves well to photometric redshift (z_p) estimation in the absence of spectroscopic redshifts (z_s): the prominent 4000Å break serves as a strong redshift discriminator in evolved galaxies. Approximately 32% of clusters in the catalogue have at least one member that has a spectroscopic redshift, while the remainder of members have photometric redshifts. Since the Stripe 82 multi-epoch data is deeper than the remainder of the SDSS DR7, some galaxies did not have pre-computed photometric redshifts. To this end, we used the hyperz utility (Bolzonella et al., 2000) to estimate redshifts for any galaxy without an existing DR7 photometric redshift or spectroscopic redshift, exploiting the deeper *ugriz* Stripe 82 photometry. As discussed in §3.2.2, the dispersion $\delta_z(z_s) = \Delta z/(1+z_s)$ for $z_p^{\rm hyperz}$ in a spectroscopically confirmed sample of 1549 galaxies within our cluster catalogue is σ_{δ_z}=0.029, compared to σ_{δ_z}=0.017 for the same galaxies when the photometric redshift is calculated with the DR7 algorithm (both figures calculated from the standard deviation in δ_z after rejecting galaxies with $>3\sigma$ clipping). We attribute the higher precision in the DR7 photometric redshifts as due to the sophistication of the DR7 approach (see Abazajian et al., 2009, for the key features

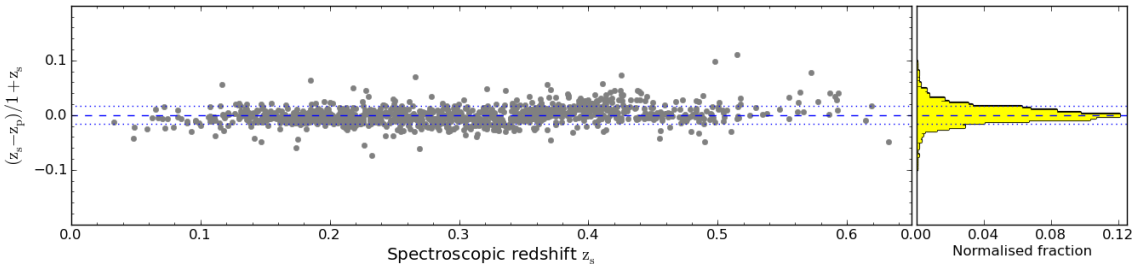

Figure 4.4: Comparison of photometric redshift accuracy $\delta_z(z_s) = (z_s - z_p)/1 + z_s$ for the brightest cluster galaxy, including both photometric redshifts derived from the DR7 database and `hyperz` estimates. After outlier rejection (clipping galaxies with $|\delta_z| > 3\sigma_{\delta_z}$, or 0.4% of the total sample), we find a 1-σ scatter σ_{δ_z}=0.016 (indicated by the blue dotted lines). This highlights the excellent redshift recovery using *ugriz* photometry alone. For a given cluster we combine both the photometric and (when available) spectroscopic redshifts of cluster members to derive a robust redshift estimate for the system as a whole.

of this algorithm) compared to our simple `hyperz` χ^2 fits to a limited range of spectral templates. For the more precise algorithm, we expect photometric redshifts to be accurate to $\sim$5–9% at the median redshift of the cluster sample. In Figure 4.4, we compare photometric and spectroscopic redshifts for the 1137 cluster BCGs with spectra. Because the resultant dispersion (σ_{δ_z}=0.016) is small, this suggests photometric redshift estimates of cluster BCGs are both accurate and a good estimate of the cluster redshift. Since we have several different redshifts for a given cluster, we can combine this information into a single redshift estimate for the cluster ensemble. As before, we calculate a weighted median redshift for the system; spectroscopic redshifts are given a weighting of 4 (in effect, that galaxy is counted four times); photometric redshifts from the DR7 catalogue have a weighting of 2, and the `hyperz` calculated redshifts are given a weighting of unity. The lower weighting for the latter reflects the slightly poorer performance of these photometric redshifts compared to those from DR7 described above. This weighting also mitigates any impact the $z_p, z_s \simeq 0.3$ "hole" noted in Figure 3.2 may have on redshift estimates: catastrophic failures in the `hyperz` estimate will not influence cluster redshifts as much, for example, errant spectroscopic redshifts. A histogram of the redshift distribution is shown in the *top-left* panel of Figure 4.5. The median redshift of clusters in the survey is

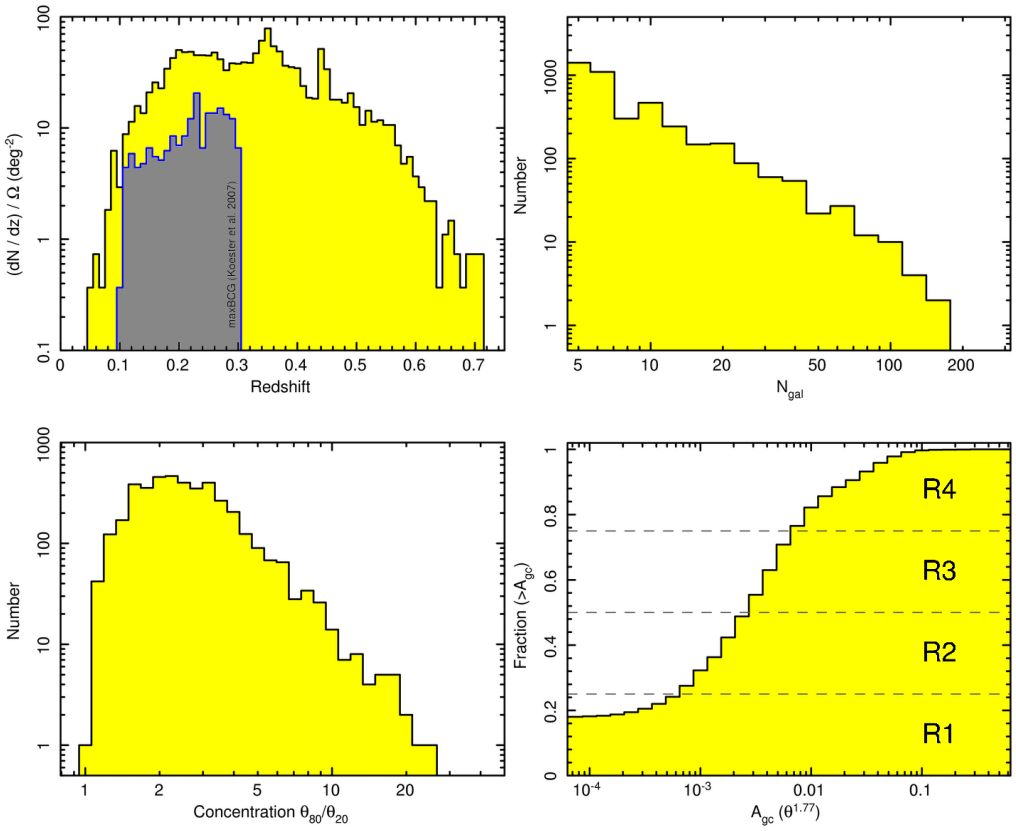

Figure 4.5: (adapted from Geach et al., 2011). Characteristics of the Stripe 82 ORCA catalogue. (*top left*) Redshift distribution for the Stripe 82 clusters compared to the distribution of clusters identified using the maxBCG cluster detection algorithm (*grey*; Koester et al., 2007a). Note that the maxBCG detector was not applied to the deeper multi-epoch photometry exploited in this work. The median redshift of clusters in the present catalogue is $z_{\rm med}$=0.32. (*top right*) The distribution of cluster membership. (*bottom left*) Concentration distribution, C, defined as the ratio of the radii of apertures containing 80% and 20% of the members. This quantity could be a useful parameter for the selection of cluster sub-sets. (*bottom right*) Richness classification (see §4.4.2), based on quartile ranges of the distribution of $A_{\rm gc}$ – the estimated amplitude of the angular correlation function of galaxies in clusters. We define four simple classes of ascending richness: R1 to R4.

z_{med}=0.32, however the depth of the multi-epoch SDSS data in this region allows us to detect clusters comfortably out to $z \sim 0.5$, with a handful of systems detected at $z \gtrsim 0.6$.

Not all survey fields have the level of spectroscopic coverage that Stripe 82 enjoys. In particularly large fields, the photometry of faint distant objects may be heavily dominated by noise, reducing the efficacy of photometric redshift estimates. Using the 4000 clusters detailed here, one may provide a more empirical route to estimating cluster redshifts, in turn unlocking many key redshift-derived cluster properties by connecting the observed ridgeline characteristics to redshift. This is illustrated in Figure 4.6 where we plot the cluster colour-colour track coded by cluster redshift. To estimate the sequence normalisation (c_{m20}) for each cluster, we perform a linear least-squares fit to the sequences identified in the two colours. Clear trends are apparent both between the colours, and the cluster redshift relative to the track they form. We emphasise this distribution is the result of accurate cluster detections rather than any input constraints; detections do not by design lie on this track, nor are they excluded from other regions of the plot.

A lookup table to these data essentially determines photometrically-estimated cluster redshifts as a side-product of the detection itself. This would be of great use in surveys with similar SDSS photometry (e.g. the Pan-STARRS 3π survey or SkyMapper[6]), where spectroscopic data will be sparse. Further calibrations against more complete spectroscopic datasets will improve the accuracy of this approach. It is significant to note that, calibrated with a large spectroscopic campaign such as SDSS3:BOSS, this approach can provide redshift estimates from cluster photometry without the assumption of any SED. Working from the same training set, it is of great interest to compare traditional photoz estimates to those derived by this method. Despite photometric redshift algorithms constraining galaxy redshifts through searches for (among other features) the 4000Å break, they require additional (and optimally many) bands to constrain the SED. Because that break only requires information in two bands (three for the method above), one may in principle arrive at the same result with considerably less data and without reliance on model SEDs or photometric redshift pipelines.

[6] `http://rsaa.anu.edu.au/skymapper`

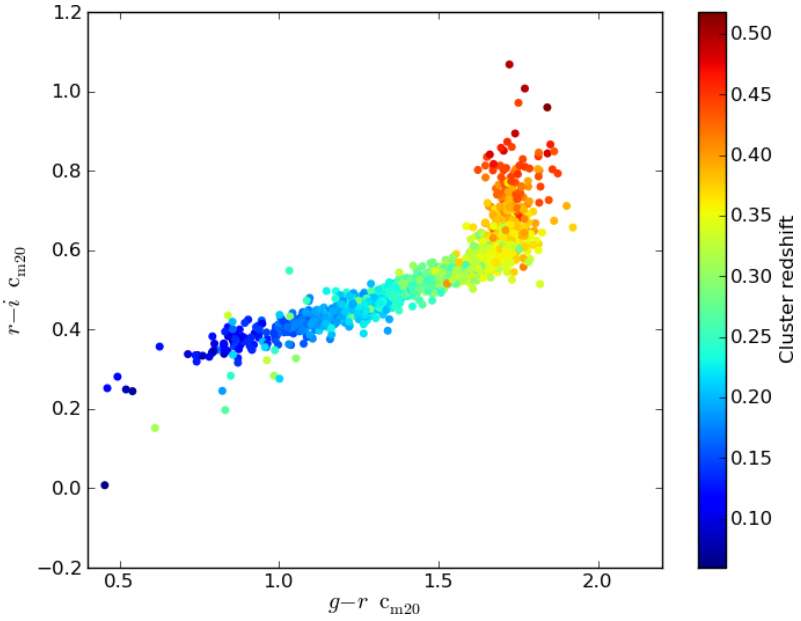

Figure 4.6: Colour-colour tracks for the ORCA 4098 clusters detected in Stripe 82. The colour normalisations in *g-r* and *r-i* are calculated for each cluster with a linear fit to the cluster red sequence. Each point is colour-coded to the calculated cluster redshift. As clusters appear redder in these bands, their estimated redshift increases.

4.4.2 Richness estimates

In light of the larger cluster population detected in this chapter, we take a closer look at the calculation of cluster richnesses described in §3.4.3.

Ideally, we should classify clusters according to an observable parameter that correlates with the mass of the structure. In the absence of X-ray luminosities, velocity dispersions or accurate lens models of the underlying matter profile, we must rely on cruder methods of richness estimation that employ counting statistics to assess the significance of the density enhancement in the cluster compared to the field. Unfortunately, calibrating optical richness measurements to various mass estimates across different surveys is notoriously difficult, and so in this catalogue we have provided several (related) richness estimates based on aperture counts of cluster members corrected for field contamination. Moreover, the information provided in the cluster catalogue should be sufficient for either the re-calculation of a specific richness estimate, or the re-calibration of our measured values to some other scale of choice.

All of our richness estimates are based on the net counts of galaxies within an aperture of radius θ. This may be centred on the interpolated cluster centre or the position of the brightest cluster galaxy (these are often to all intents the same). The net counts are

$$N_{\rm net} = N_T - N_B, \tag{4.1}$$

where N_T is the number of galaxies within the aperture and N_B is the number of background galaxies inside an annulus centred on the cluster, with equivalent area to the N_T selection aperture. Although using an annulus instead of a scaled surface density for the full catalogue results in poorer number statistics, it does account for potential differences in photometric properties (such as seeing, local extinction) across different regions of the stripe. In this prescription, we have made no correction for the presence of bright stellar haloes, or other cosmetic effects that might impact the aperture counts. We adopt two values for the aperture scale θ:

1. θ_{80} - the radius of an aperture containing 80% of the members.

2. $\theta_{0.5{\rm Mpc}}$ - the angular size of an aperture with projected physical size $0.5h^{-1}{\rm Mpc}$.

The latter is based on our estimates of cluster redshift. The galaxy counts N_T and N_B can be defined as either:

i. All galaxies in the magnitude range $[m_3, m_3 + 3]$.[7]

ii. All galaxies in the photometric filter the cluster was detected in.

There are issues with both counting schemes that introduce uncertainties in the richness calculation. We chose m_3 as a counting reference because it is purely empirical and can easily be derived from the catalogue without any additional assumptions about the cluster luminosity function. However, the scatter in m_3 will inflate both for low-membership and high-redshift clusters due to stochasticity, photometric uncertainties and projection effects. For example, in a redshift slice $0.2 < {\rm z} < 0.3$ the standard deviation of m_3 measured for all clusters is strongly dependent on the number of galaxies assigned to the cluster, $N_{\rm gal}$. For clusters with $5 \leq N_{\rm gal} \leq 10$ we find $\sigma_{m3} = 0.77$, dropping to $\sigma_{m3} = 0.37$ for richer systems between $25 \leq N_{\rm gal} \leq 35$. Similarly, counting

[7] where m_3 is the magnitude of the third brightest cluster member.

galaxies only in the photometric filter will result in uncertainty due to galaxies being scattered out of and into the slice - an issue that is also exacerbated for low-mass/faint systems.

Despite their limitations, from these basic statistics we can derive higher order richness estimates. One commonly used measure is the $B_{\rm gc}$ statistic (Longair and Seldner, 1979; Yee and López-Cruz, 1999) which has been shown to scale well with other more direct measurements of cluster mass (Yee and Ellingson, 2003). This statistic is designed to estimate the amplitude of the spatial cross-correlation function for galaxies:

$$\xi(r) = B_{\rm gc} r^{-\gamma}. \tag{4.2}$$

To calculate $B_{\rm gc}$ requires a de-projection of the amplitude of the *angular* correlation function into 3D space, and this is estimated by:

$$B_{\rm gc} = \frac{\rho_g A_{\rm gc}}{\Phi(m_l, {\rm z}) I_\gamma} d_\theta^{\gamma-3}, \tag{4.3}$$

where $\rho_{\rm g}$ is the average surface density of galaxies in the field brighter than a limit m_l, d_θ is the angular diameter distance to the redshift of the cluster, γ is the slope of the power-law in the correlation function (Equation 4.2) and $A_{\rm gc}$ is the amplitude of the angular correlation function. $A_{\rm gc}$ is in turn estimated as:

$$A_{\rm gc} = \frac{N_{\rm net}}{N_B} \frac{3-\gamma}{2} \theta^{\gamma-1}. \tag{4.4}$$

The $B_{\rm gc}$ statistic is finally scaled by the luminosity function $\Phi(m_l, {\rm z})$, integrated between the absolute magnitude of the second brightest cluster member, down to a luminosity corresponding to m_l at the cluster redshift. We characterise the evolution of the luminosity function following the approach of Yee and López-Cruz (1999) by adding a small redshift-dependent term to $M^\star$. The integration constant is set a value $I_\gamma = 3.78$, we use $\gamma = 1.77$, and the limiting magnitude $m_l = m_3 + 3$. $B_{\rm gc}$ and $A_{\rm gc}$ for the variants of angular aperture (1. and 2.) and counting method (i. and ii.) are provided in the cluster catalogue. For the richness classifications in the cluster catalogue, we prefer the angular correlation function amplitude $A_{\rm gc[1,i]}$ calculated for θ_{80} and galaxies selected in the set of photometric filters the cluster was detected in. Unlike $B_{\rm gc}$, this requires no scaling for luminosity function. We emphasise that all of these statistics are ultimately governed by Poisson noise in N_T and N_B - a break-down of practicality in these richness statistics for low-member (group) systems should be expected.

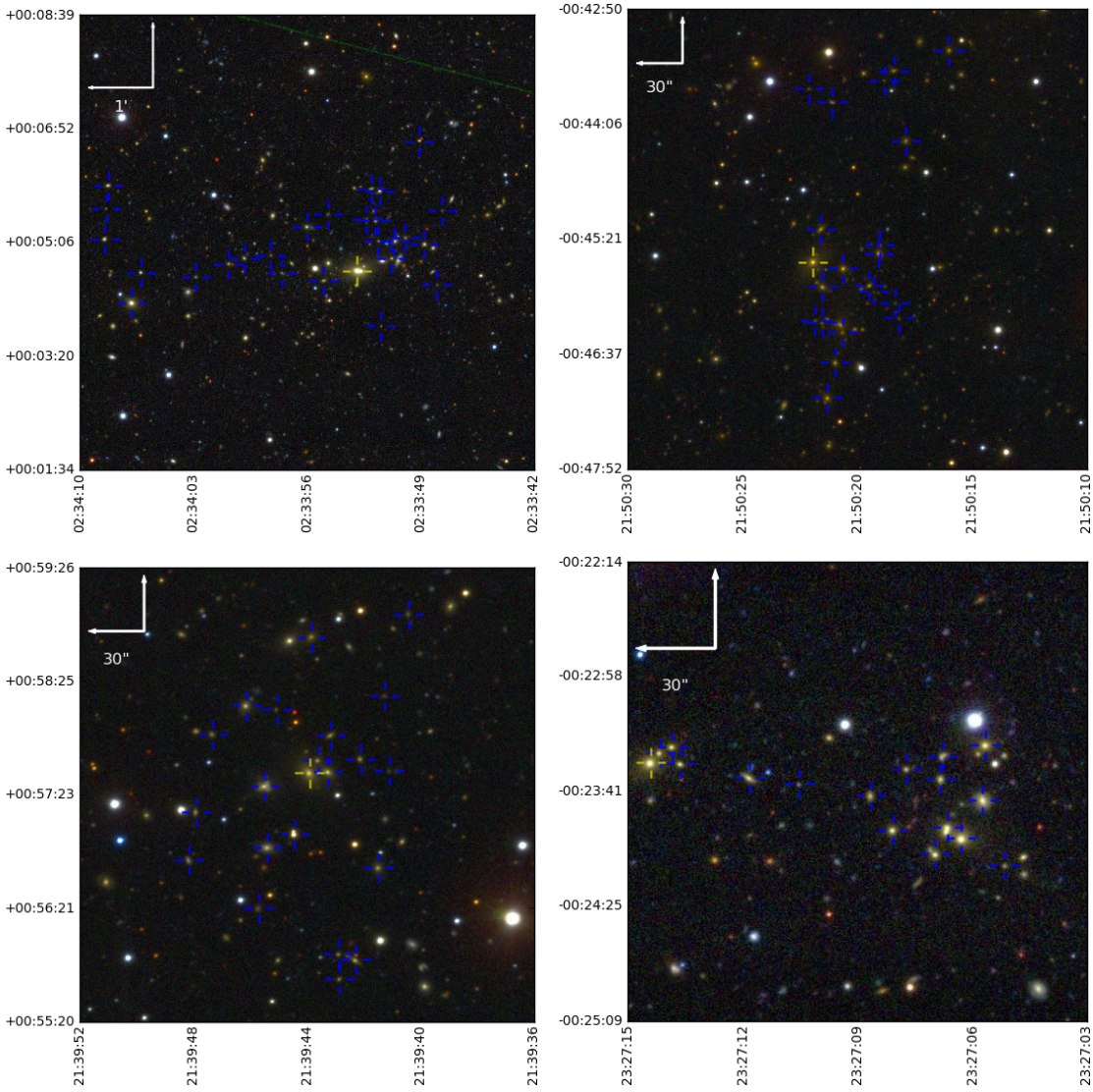

Figure 4.7: Rich `ORCA` clusters without `maxBCG` counterparts. From *top-left* to *bottom-right*, these clusters are *GMB J023356+00050.1*, *GMB J215021-00450.4*, *GMB J213944+00570.4* and *GMB J232709-00230.7*. All are class R=4 richness clusters with redshifts of 0.18, 0.28, 0.19 and 0.29 respectively.

In order to coarsely segregate the catalogue into richness bins, we define four classifications of richness: R1 to R4. These classifications are simply the quartile ranges of the parameter $A_{\mathrm{gc[1,i]}}$; the cumulative histogram and ranges are given in the *bottom-right* panel of Figure 4.5. In combination with a richness estimate, the concentration of galaxies in a cluster can also be a useful parameter to describe the morphology of the system.

We define a simple dimensionless concentration parameter $C = (\theta_{80}/\theta_{20})$, where θ_{20} and θ_{80} are the radii of a circular aperture containing 20% and 80% of the members respectively. We find a median concentration of $\langle C \rangle \sim 2.4$ (the *bottom-left* panel, Figure 4.5).

In Figure 4.7, we highlight four example clusters in the richest (R4) category that do not have `maxBCG` counterparts. These detections were selected from clusters between $0.15 < \mathrm{z} < 0.3$ (i.e within the `maxBCG` detection range), separated by at least $30'$ from the declination limits of the survey (to preclude any impact edge effects may play in either algorithm), but with at least 1° degree separation from the closest `maxBCG` cluster. The investigation of such systems will be a useful insight into the differences in cluster selection function between `ORCA` and `maxBCG` .

4.5 Revisiting the blue limit

In §3.3.7, we defined blue limits in the bluest colour pair (*g-r*) to both reduce the parameter space searched and to prevent the detection of sequences bluer than the Virgo cluster (estimated as $c_{\mathrm{m20}} = 0.47$). To establish how justified we are in applying this limit, `ORCA` was re-applied to the whole stripe to search for sequences with *g-r*<0.47. As with the previous studies, a secondary sequence detection of all systems in *r-i* was required. Over *g-r*<0.47, we identify a total of 41 cluster candidates covering 270 square degrees. Following the visual inspection of these systems, at least 23 were clustered fields of stellar systems with very uniform colours - it is possible these are open clusters local to the Milky Way (for an example, see the *left* panel of Figure 4.8). In Figure 4.9, we superimpose the 41 detected systems (*red* and *blue* points) onto the *g-r* vs *r-i* sequence normalisation track presented in Figure 4.6. The trend-line is included to visually aid the extrapolation of the colour-colour relation beyond the limits of the cluster catalogue. *Blue* points indicate the "stellar clusters", *red* are detections with galaxies. The two populations appear well separated, with most of the stellar detections far from the extrapolated cluster track. Assuming only clusters lie on this relation, an additional 12 new systems close to the extrapolated track (∼0.3% of the full cluster catalogue) could be included. Closer inspection of these detections reveal predominantly faint, blue galaxies with large photometric errors. In the *right* panel of Figure 4.8 we show one of the more promising candidates, lying directly on the extrapolated track with three others. This

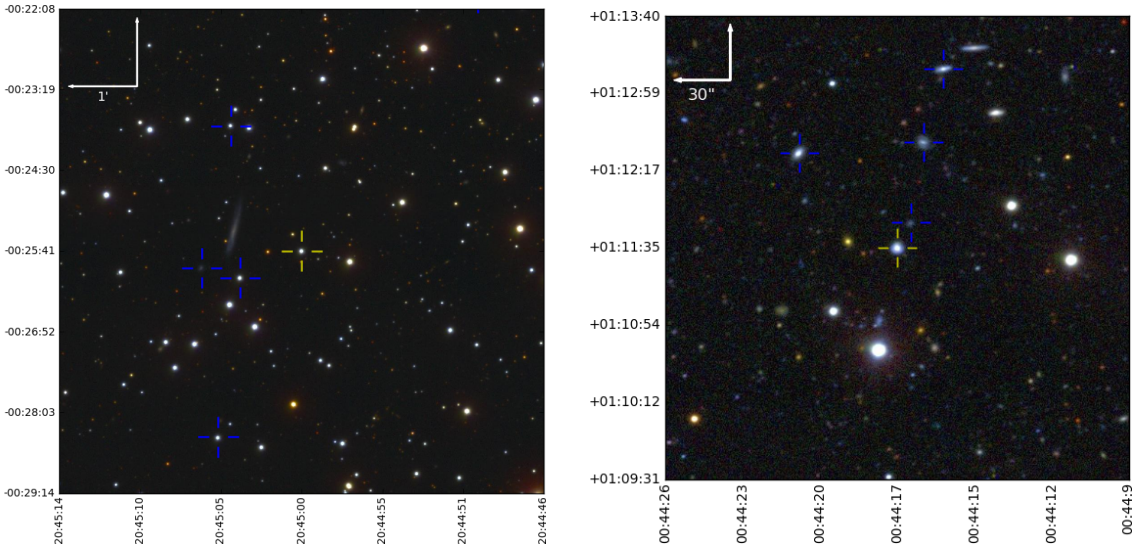

Figure 4.8: Two systems identified by ORCA when searching for clusters below the blue *g-r* limit. The *left* example is a detection in a stellar field, whilst the system depicted on *right* consists of galactic sources.

system has one spectroscopically confirmed redshift of z=0.1, and photometric redshift estimates consistent with this measurement. However, also noted is one stellar system situated close to the track. Given only the modest number of extra clusters that would be included, doubts over their veracity, and the proximity of some stellar sources to the extrapolated track, setting a g-r>0.47 constraint ensures ORCA does not include potentially false detections.

For reference, we include the positions of the Coma and Virgo clusters on this plot as the black and white stars respectively. These points were derived from extrapolations of the colour-magnitude relation down to $20^{\rm th}$ magnitude. We produced the Coma CMR with Sloan photometry from SDSS object IDs (John Lucey, priv. communication). Virgo positions were cross-matched between the Virgo GOLDMine database (Gavazzi et al., 2003) and SDSS data. We note that our improved value for the g-r $c_{\rm m20}$ Virgo normalisation is 0.16 magnitudes redder than the 0.47 estimated from the Rines and Geller (2008) study. This discrepancy may arise from the deeper sample we extract from the GOLDMine data (0.5 magnitudes) and our choice of model rather than fibre magnitudes when calculating galaxy colours.

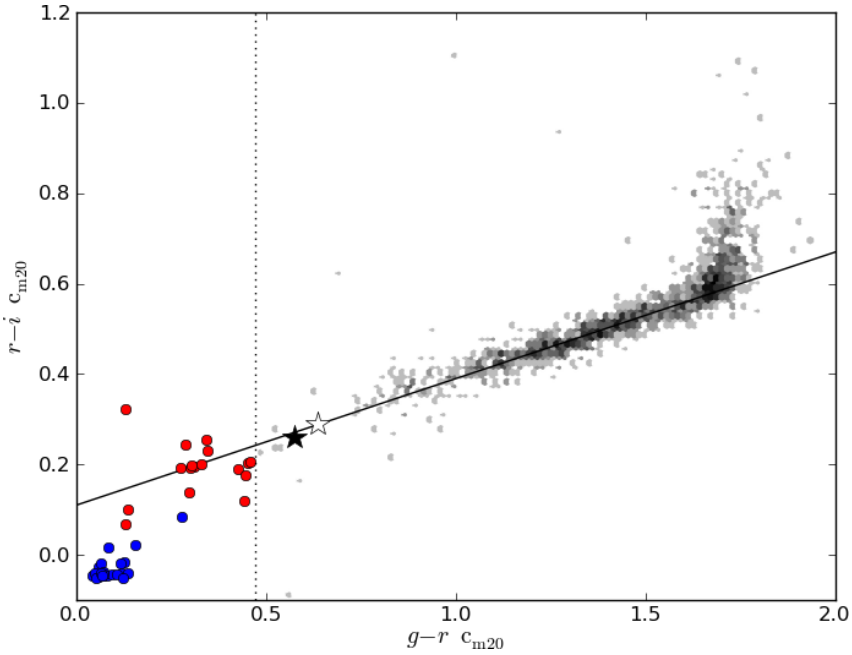

Figure 4.9: The colour properties of systems detected by ORCA below the *g-r* blue limit. The grey-scale data indicate *g-r* and *r-i* sequence normalisations of the 4098 Stripe 82 cluster sequences (as per Figure 4.6), with logarithmically-spaced intensity indicating the number of clusters. Blue points are visually identified stellar clusters, red points are detections containing galaxies. The black and white star symbols indicate the positions of Coma and Virgo respectively. The *black* vertical dotted line denotes the blue *g-r* limit set in chapter 3.

4.6 Redshift evolution of the CMR slope

As observed in both real (e.g. Figure 3.10) and mock (e.g. Figure 3.12) cluster sequences, bright galaxies appear redder than their fainter companions. This effect, manifesting itself as a slope in the colour-magnitude relation, may be used to infer the properties of the cluster. Due to the age-metallicity degeneracy (see e.g. Worthey, Trager, and Faber, 1995), the actual underlying mechanism behind the slope is controversial. Fainter galaxies in the sequence are either metal deficient or have younger stellar populations.

Kodama and Arimoto (1997) broke the degeneracy by comparing age-driven and metallicity-driven sequence evolution of two high-redshift ($z = 0.23, 0.4$) clusters. Their analysis suggested a mass-metallicity (observed as magnitude-metallicity) relation drives

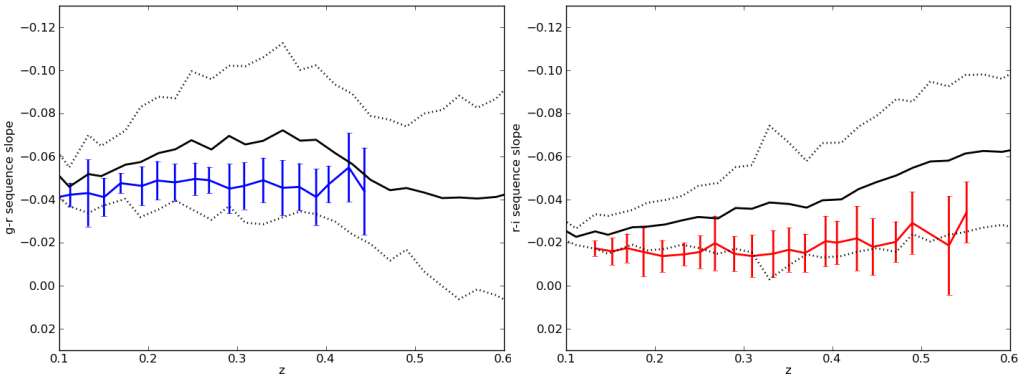

Figure 4.10: The redshift evolution of the red-sequence slope in *g-r* -selected clusters (*left*) and *r-i* -selected clusters (*right*). The solid black line denotes sequence slopes from the models, with dotted lines indicating $\pm 1\sigma$ uncertainties.

the sequence, arising from supernova feedback ejecting gas more efficiently in smaller galaxies. That the observed sequence fails to evolve significantly over the cluster look-back times further disproved the age-driven scenario.

The redshift range we probe here provides a useful verification of previous CMR slope studies, but with many more clusters. We are also able to determine how well matched the models are to our observed slopes. Exploiting the accurate cluster redshift estimates described in §4.4.1, we compare observed-frame ORCA red-sequence slopes to those derived from a mock galaxy catalogue. This catalogue (A. Merson, priv. communication) consists of *griz* photometry, angular, redshift and stellar mass data for galaxies covering 100 square degrees. We construct a mock-cluster sample from this mock catalogue by identifying all galaxies residing in dark matter haloes with at least four other galaxies. We restrict our study to clusters with $z \leq 0.7$ and total stellar masses exceeding $10^{9.5} h^{-1}\,\mathrm{M}_{\odot}$. To determine the sequence slopes, we compute a linear least-squares fit to the colour-magnitude galaxy data for each mock cluster.

To select high-probability detections in the ORCA dataset we only analysed clusters with at least 10 members. We split this subset into two populations: those detected primarily in *g-r*, and those in *r-i* (although we note here the former case were secondarily identified in *r-i*). For each of the clusters in these two groups (593 and 441 respectively) we perform the same fitting procedure to derive their red sequence properties. In Figure 4.10 we show the redshift evolution of the observed-frame ORCA slopes compared to the

model. In both colours, this slope is systematically shallower than the model over all redshifts, and only in *r-i* does it appear to show a modest evolution. We find this in agreement with the observed-frame *V-I* slope evolution measured in Stott et al. (2009), where the gradient begins to vary beyond $z \sim 0.4$. The lack of evidence for evolution over the redshift range we measure supports other studies in suggesting the underlying mechanism for the sequence slope is not driven by age.

Of particular interest is the systematic steepening of sequence slopes for model clusters; the role galaxy formation models play in this effect is unclear, and further complicated by the K-correction. In the Bower et al. (2006) model used in this study, the predicted satellite (i.e. lower luminosity) galaxy colours are redder than those observed in real clusters, arising mainly from the rapid removal of cold gas in simulations. This effect should serve to flatten the sequence slope. Efforts to improve the transition of galaxies onto the red sequence (e.g. ram-pressure stripping; McCarthy et al., 2008) will only serve to further steepen the slopes of these cluster ridgelines. Whilst it remains a challenge to correctly reproduce the correct galaxy cluster colours, performing similar analyses with alternative galaxy formation models can help to constrain the parameters describing the physics responsible for the slope.

4.7 Cluster BCG colour tracks

With accurate photometry and redshift estimates for our clusters, we may also compare the photometric properties of cluster galaxies to model data. In Figure 4.11 we compare the colour (*g-r*, *r-i*, *i-z*) evolution of brightest cluster galaxies in Stripe 82 to predictions from the latest composite stellar population models designed to replicate the broad band colours of Luminous Red Galaxies (Maraston et al., 2009). In this plot we use clusters with at least 10 members (1100 in total) and the cluster redshift which, as demonstrated in Figure 4.4, provides an accurate estimate of the BCG redshift. We emphasise here that galaxies were not selected to lie on these relations, but instead follow well described tracks corresponding to red, quiescent stellar populations. As discussed in §4.4.1, the tight colour sequence allows one to predict photometric redshifts for the clusters with high accuracy. In the absence of spectroscopic redshifts, we are therefore confident of all cluster redshifts in our Stripe 82 sample. The *g-r* track most sensitive to the 4000Å break

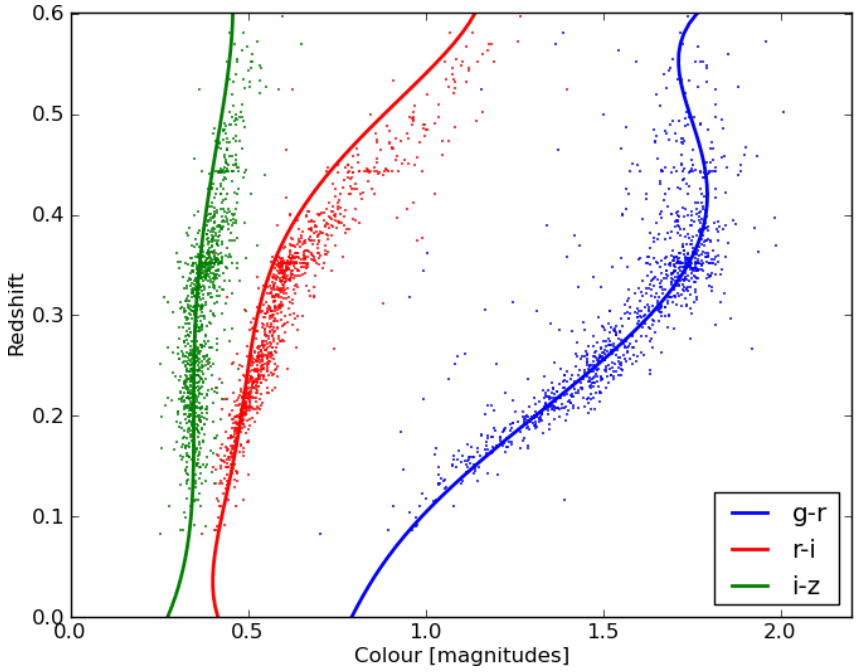

Figure 4.11: The redshift evolution of ORCA cluster BCG colours compared to those predicted by the Maraston et al. (2009) LRG model.

out to z $\sim$0.4 agrees well with the data, but both the *r*-*i* and *i*-*z* model data are systematically bluer than the ORCA BCGs.

To quantify the agreement between model and data, we determine the distribution of BCG colours about the *g*-*r* model track. A median colour shift $\Delta(g-r)_{\rm m}$ of -0.005 was calculated (negative indicating the data are bluer than the model), along with a 1-σ scatter of $\Delta(g-r)_{\sigma}$ $\sim$0.2. We next investigate how cluster member colours shift relative to model expectations as we probe down the sequence in galaxy luminosity. According to the (galactic stellar-)mass-metallicity relation from which the sequence arises, we should find less luminous galaxies, having lower metallicity, are bluer. In Figure 4.12, we plot the model-adjusted distribution of colours for L$^{\rm th}$-ranked luminosity galaxies in ORCA clusters containing $N_{\rm gal}$ $\geq$10, for L=1 (BCG) to L=6. The *bottom*-panel histograms of *g*-*r* colours show a progressive blue-ward shift as we probe fainter down the sequence, but with the colour scatter largely constant with L. The redshifts used for this analysis are for the clusters themselves. Both Figure 4.11 and Figure 4.12 show a slight drop in cluster number at z $\simeq$0.3. Although this incompleteness does not influence the findings we present here, it is perhaps indicative of the hyperz algorithm failure at this redshift (seen

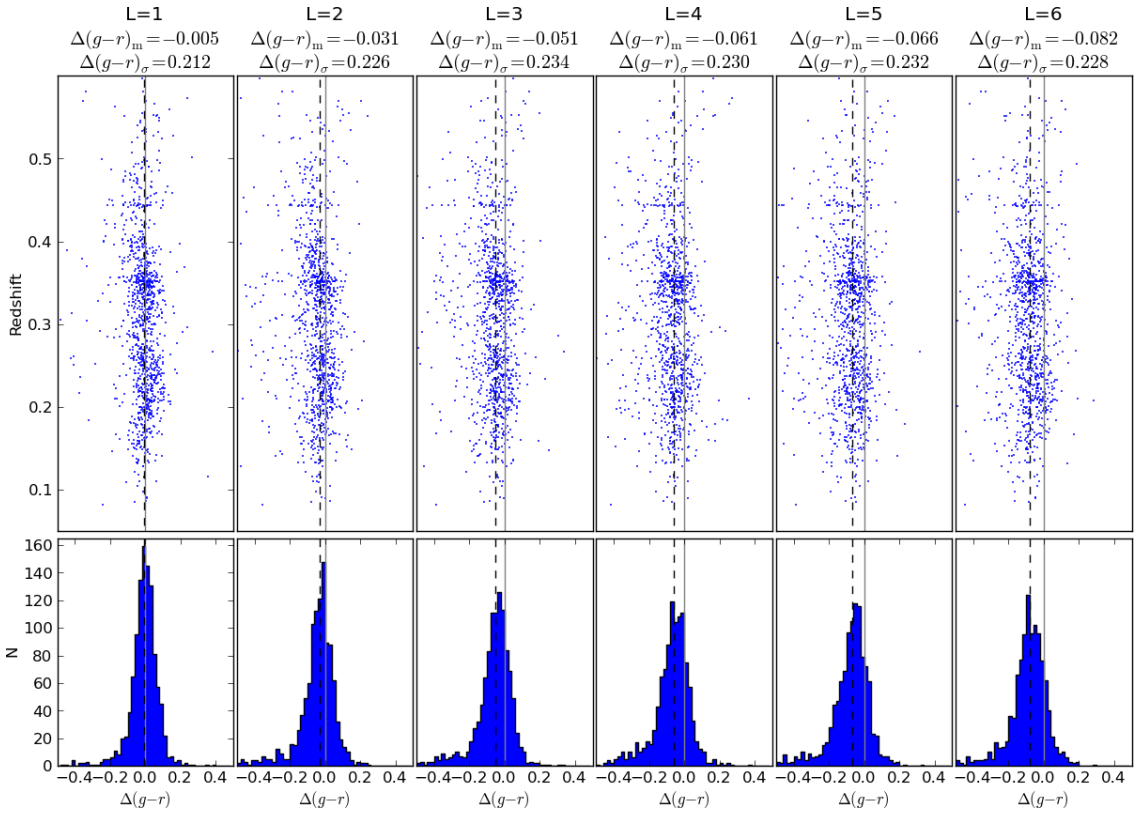

Figure 4.12: The *g-r* colour distribution of the L^{th} luminosity-ranked galaxy compared to the Maraston et al. (2009) LRG model. As one probes to lower masses, the median galaxy colour shifts bluer. The black dotted line, $\Delta(g-r)_m$, indicates the shift in colour with respect to the model, and $\Delta(g-r)_\sigma$ is the standard deviation of the scatter about this median.

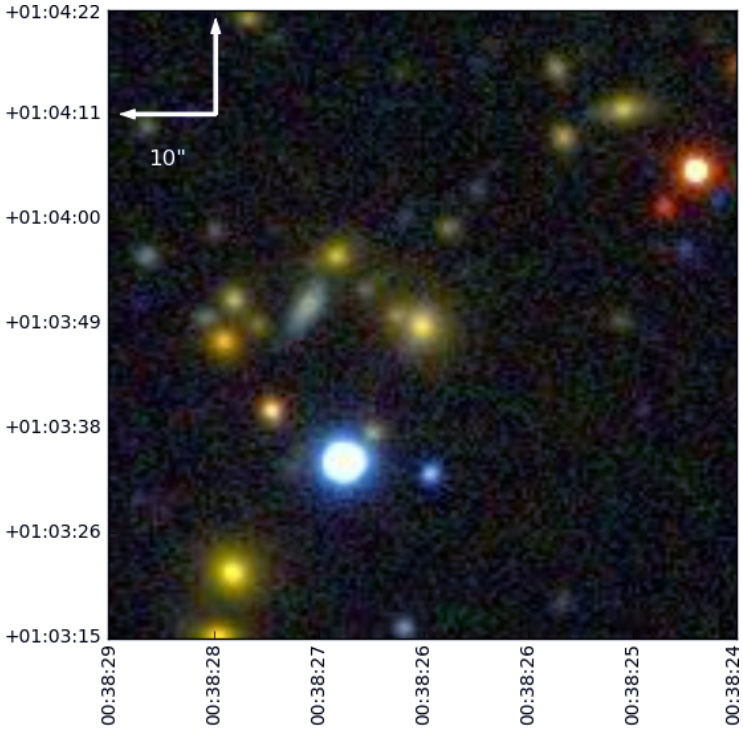

Figure 4.13: A *gri*-composite image showing the BCG of cluster *GMB J003827+01030.9*, a z=0.29 cluster with 10 members. This BCG was the 17^{th} bluest BCG amongst our sample of 1100 with respect to the reference model.

in Figure 3.2 and discussed in §4.4.1). Although the spectroscopic sampling rate drops considerably with increasing L, when re-analysing these data with either spectroscopic or photometric redshifts of the galaxies themselves, the progressive bluewards migration is still observed, albeit slightly attenuated.

One explanation for the observed trend is an age-driven mechanism: lower luminosity galaxies have experienced more recent episodes of star formation. Although producing similar blue-ward shifts in galaxy colour, such an effect would also introduce a tilt to the scatter plots in Figure 4.12. Because these data are colinear with the redshift axis, we instead interpret this trend as the mass-metallicity relation, and in effect here we probe the slope of the CMR.

Although few in number, some cluster BCGs are noted to be significantly blue. It is possible the cluster cores of these systems either show evidence of low-level star formation (see e.g Bildfell et al., 2008) or are host to an AGN. Spectroscopy with the

SDSS3:BOSS instrument will target BCGs such as the galaxy centred in Figure 4.13 (the $17^{\rm th}$ bluest BCG in the subset investigated) and provide more insight into their nature.

4.8 Summary

In this chapter, we have highlighted some of the developmental aspects required to feasibly search for clusters in deep and wide photometric surveys. Our modifications to ORCA allow for faster computational processing, but still requires some partitioning of the data.

We apply our updated algorithm to the Sloan Digital Sky Survey "Stripe 82" equatorial multi-epoch co-add, a 270 square degree stripe with photometry approximately 2 mags deeper than the general SDSS imaging survey. From this survey we identify 4,098 clusters, consisting of $\sim$60,000 cluster galaxies with a cluster median redshift $z_{\rm med}$=0.32. With access to redder SDSS photometric bands, ORCA comfortably detects systems out to $z \sim 0.6$. For each detected system, we derive cluster richness statistics based on counts of galaxies relative to the background. The Stripe 82 cluster catalogue is publicly available, and will be maintained online[8]. Full details on the catalogue content are provided in Appendix C. We expect to improve the catalogue in future releases, with follow-up imaging, spectroscopy, and refinements of the detection algorithm.

We verify that our selection of the blue limit, tied to the red sequence of the Virgo cluster, is appropriately chosen. When detecting candidate ORCA clusters below this limit, we identify a number of stellar clusters. Inclusion of the identified systems would only increase the sample size by 0.2%, while potentially adding false detections to the catalogue. Using accurate cluster redshifts and photometry, we compare the derived sequence slopes of our cluster sample to ΛCDM models. The reconstruction of mock clusters from a lightcone simulation permits comparison of the ORCA red sequence slope redshift evolution to model predictions. There is a broad agreement between the two, with evidence of only mild evolution in ORCA data over the redshift range studied. Finally, we compare the ORCA BCG colour evolution to predictions from a stellar population LRG model. Model and ORCA colour tracks agree well for the *g*-*r* colour, but fainter ORCA members become progressively bluer with respect to the model. In effect probing

[8] http://astro.dur.ac.uk/$\sim$dmurphy/orca/data

the red sequence slope, we exclude stellar age as the underlying mechanism, instead attributing this effect to the mass-metallicity relation. We do not, however, exclude the possibility of low-level star formation in some cluster members.

Chapter 5

Applying and developing ORCA

5.1 Overview

This chapter contains a selection of the surveys ORCA has been applied to, and provides a snapshot of the work currently in progress. Data from these studies have either been used in projects that will be published, or (in the case of the penultimate section) will form part of a paper themselves. These investigations also highlight some of the additional algorithm development required to correctly analyse the types of survey data routinely encountered. With access to Pan-STARRS y-band photometry, we demonstrate how one may estimate cluster stellar masses, and relate them to observables produced from ORCA cluster catalogues. The largest cluster catalogue we discuss here, derived from SDSS DR7 photometry covering $\sim$8,000deg^2, demonstrates the cluster detection pipeline is well-placed to process data on the scale of the forthcoming Pan-STARRS 3π survey. Table 5.1 provides an overview of the cluster catalogues produced from work detailed in this chapter; this is a non-exhaustive list of the surveys ORCA has thus-far been applied to. Before analysing these catalogues, we introduce added functionality to the ORCA detector that permits exploration of a wider range of surveys. Following this, we investigate the surveys in the order listed in Table 5.1.

5.2 Edge detection and hole identification

In this section, we discuss the sensitivity of the ORCA detector to surveys with irregular boundaries. For illustrative purposes, in the *first* panel of Figure 5.1 we show the positions of galaxies from an example survey we investigate later in §5.7, with a series

Section	Survey	Area (deg^2)	$N_{galaxies}$	$N_{clusters}$	$N_{cl-gals}$	$\bar{N}_{gal}$	z_{median}
§5.3	PS1 MD-8	14.37	634,825	111	1,248	11.2	0.26*
§5.4	PS1 SAS	61.78	649,668	153	1,233	8.1	0.21
§5.4	SDSS SAS	61.78	647,789	211	1,934	9.2	0.22
§5.5	PS1 MD-5	11.59	919,621	133	1,342	10.0	0.29*
§5.6	SDSS-CR	7621	23,829,607	32,808	335,581	10.2	0.21
§5.7	DXS/SA22	5.3	303,008	9	54	6.0	0.87

Table 5.1: The properties of surveys featured in this chapter. The final five columns indicate the number of galaxies in the survey, the number of clusters detected and how many galaxies were assigned to them, the mean number of galaxies per cluster and the median redshift of clusters in the survey.

* Redshift data for this catalogue are incomplete

of circular regions excised of galaxies. In this instance, they are due to the removal of stellar sources from the field. More generally, voids in data will also occur when, in a wide-area scan, the survey did not cover that portion of the sky. We will next explain the two reasons why ORCA must be able to account for these survey features, and then return to this Figure in our description of how it does so.

Firstly, where a significant fraction of the survey field is empty, the "filling factor" of the data is reduced, and simplified calculation of the mean galaxy density (the number of galaxies measured divided by the area) no longer applies. ORCA requires an accurate estimate of the mean density: an underestimate (as is implied with filling factors $<$1) in effect relaxes the P_{thresh} constraint, resulting in the amplification and subsequent selection of low significance spurious sources.

We illustrate this in Figure 5.2, where the two panels show ORCA clusters detected from the *same* source data, but with different values of the mean density passed to ORCA. In the *right-panel* we supply ORCA with the correct mean galaxy density, but in the *left panel*, we supply ORCA with half of that value. Not only are more cluster detections generated (by a factor of $\sim$3) in this underestimated density, but the median Voronoi cell area is also larger (by a factor $\sim$1.8). This latter effect arises from the cells ORCA considers

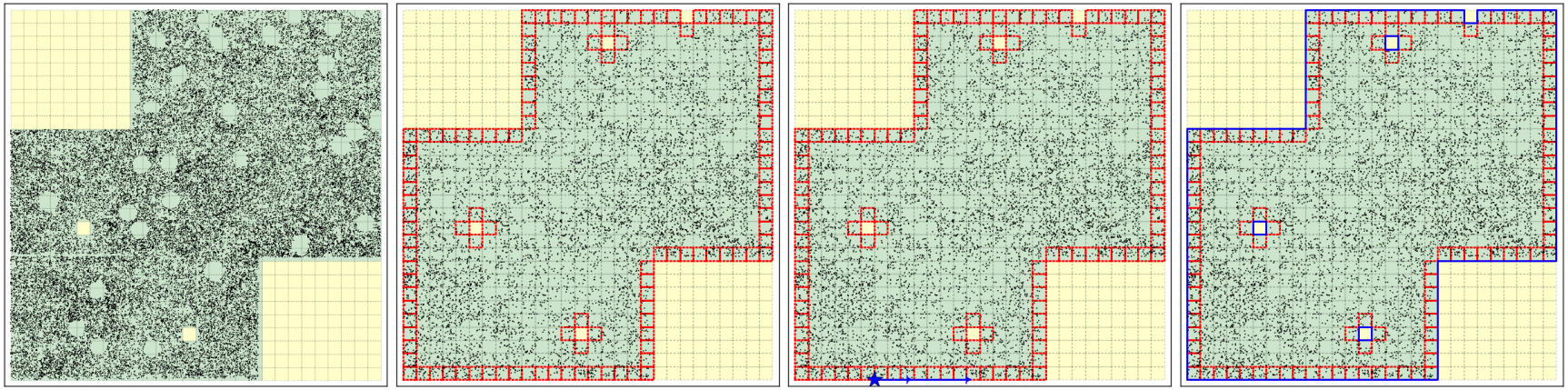

Figure 5.1: A depiction of the DARB algorithm, applied to the SA22/DXS dataset studied in §5.7. In the *first panel*, containing the angular positions of ∼45,000 galaxies, the complicated survey edge and excised regions are evident. For clarity, in the subsequent panels we have reduced the number of galaxies. The algorithm lays down a regular grid and identifies any cells containing galaxies (turquoise). In this example the cell sizes are much larger than those typically applied to the target data. In the *second panel*, occupied "edge cells" sharing at least one side with an unoccupied cell are highlighted in red. In the *third panel*, an arbitrary edge cell is selected (the blue star), and an anticlockwise walk over connecting edges is performed. This walk continues until DARB arrives back at the starting point, thus completing the delineation of one boundary. In the *fourth panel*, DARB has completed all three walks. The three smallest boundaries are classified as "holes", and the largest as an "island". Some datasets (e.g. the SDSS DR7 data studied in §5.6) have fields with multiple islands - a consequence of the gridding used to partition large contiguous surveys.

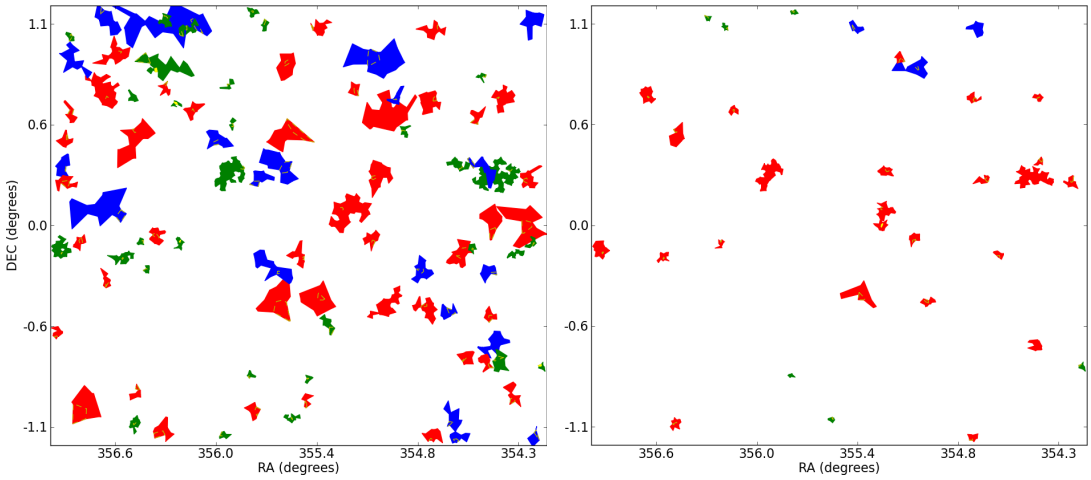

Figure 5.2: Two detection runs illustrating ORCA's dependence on the correct mean area estimation of the survey field. When the mean galaxy density is underestimated (*left*), too many objects are detected, with on average larger Voronoi cell areas. Cell colours correspond to colour pairs the clusters were identified in. Blue cells are clusters detected in the $\{g\text{-}r, r\text{-}i\}$ filter pairs, red are clusters detected in $\{r\text{-}i, i\text{-}z\}$, and green are those detected in $\{i\text{-}z\}$.

"statistically significant" relative to cell areas from a random distribution. Because the areas are normalised to the mean galaxy density when calculating $\mathrm{P(a)}$ (Equation 3.1), Voronoi cells larger than ordinarily permitted are selected.

The second reason, that ORCA must also account for irregular data, stems more from comparing detections to other datasets. If for example an X-ray cluster is located in a partially or wholly unobservable position, this must be considered when calculating the optical survey completeness relative to the other dataset.

To address these concerns, we develop an ORCA module called DARB ("Define ARbitrary Boundary"). This algorithm performs two tasks: it determines boundaries and holes in any 2-D spatially-distributed dataset and makes an estimate of the sky area covered by the data. We outline the algorithm steps here, illustrated with example data in Figure 5.1. Our first three points below are used to determine the survey area, while points 4-6 delineate distinct parts of the data:

1 The positional data is read into memory.

2 A regular grid is overlaid, with the resolution specifying the cell size.

3 Any occupied cells are added to a list. Those occupied cells sharing fewer than four sides with occupied cells are labelled as "edge cells".

4 Selecting the first occupied edge cell, an anticlockwise "walk" through connected neighbouring edge cells is performed until the algorithm arrives back at the original cell.

5 Storing this path, a new walk starts at the next edge cell yet to be visited. This is repeated until no more occupied cells remain, ensuring all distinct bounded data regions have been sampled.

6 Routes are compared and classified as "islands" (the space enclosed by the route contains galaxies) or "holes" (the space enclosed by the route contains no galaxies)

The area is estimated from the filling factor of the dataset - the fraction of cells occupied by galaxies. For the majority of surveys tackled, the approach detailed here is more than sufficient. By using the boundary data as a template, DARB also acts as a useful tool for recreating exactly the same footprint in a different dataset. This is put to use in §5.4.

5.3 Pan-STARRS Medium Deep Survey #8

The ORCA algorithm was originally designed for Pan-STARRS, serving as one provider of cluster detections from five-band photometric data. A cluster-finding effort using the "Probability Friends-of-Friends" approach (Liu et al., 2008) has been developed in tandem.

The study of extragalactic science with Pan-STARRS focuses on two types of survey, and will make use of over 1 petabyte of data from the PS-1 telescope. The Medium Deep Surveys (MDS) will be a set of 10 deep fields covering a total of 84deg^2 down to (g, r, i, z, y) 5σ point-source depths of (27.3,26.9,27.9,26.3,24.8). The second survey will comprise shallower data covering 3/4 of the visible sky - we analyse some of these data in §5.4. This section explores early imaging from the eighth field of the MDS sample (MD-8), covering the ELAIS N1 survey region and centred at (α,δ)=(246.8°, 55.0°). The 9deg^2 Pan-STARRS coverage for this field is supported by a wealth of multi-wavelength

data, including *Spitzer* SWIRE and deeper SERVS coverage, planned *Herschel* HerMES, spectroscopic data from MMT's Hectospec EN1X pointing and UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al., 2007) DXS imaging.

An MD-8 extended source catalogue was generated at Durham (P. Draper, priv. communication). The `SExtractor` (Bertin and Arnouts, 1996) utility identified sources in the *i*-band, and assigned 3″ aperture magnitudes to positions in this and remaining bands. Because the detection threshold was set to 1.25σ significance, many of the catalogue sources were spurious detections. To remove those, we set conservative magnitude limits of (g, r, i, z, y) = (24.5,24.5,24.3,23.6,22.2). These limits are established in the same manner described in §3.3.7: we filter out all galaxies faintward of the magnitude corresponding to an uncertainty Δm=0.117.

Because the fluxes across filters for these sources are unlikely to be correlated, any remaining spurious noise should be filtered out when searching for sequences in two colours. From the applied magnitude limits, we generate an input catalogue comprising 634,825 sources with angular positions and *grizy* 3″ aperture magnitude photometry. To calculate cluster redshifts after their detection with `ORCA` , we cross-match the input galaxy catalogue with the positions of SDSS galaxies having redshift data, assigning photometric or spectroscopic redshifts wherever matches to within 1″ are found. However, because the deeper photometry in this PS-1 sample makes it unlikely accurate redshift information is gained from the SDSS, any derived cluster redshifts should be treated with suspicion. Moreover, because at the time of investigation Pan-STARRS photometry was unreliable, photometric redshifts were unavailable from the PS-1 photometric redshift pipeline. We searched for cluster detections in the following filter combinations: {*g*-*r*, *r*-*i*}, {*r*-*i*, *i*-*z*} and {*i*-*y*}, adopting the filter parameters as described in the Stripe 82 analysis (Table 3.2). Because the *y*-band red-sequence is not currently well defined, we set the *i*-*y* filter slope to that of the *i*-*z* estimate, and apply the same photometric selection filter width of 0.152. This large width makes our choice of filter slope less critical.

The `DARB` algorithm was applied to the galaxy catalogue to determine the survey area (14.4deg^2) and boundary. Following a full search of this survey, the `ORCA` detector allocated 1248 galaxies into 111 clusters. In Figure 5.3 we show the position of these detections and the colours they were selected in. A table detailing the properties of each cluster can be found in §C.3 of Appendix C.

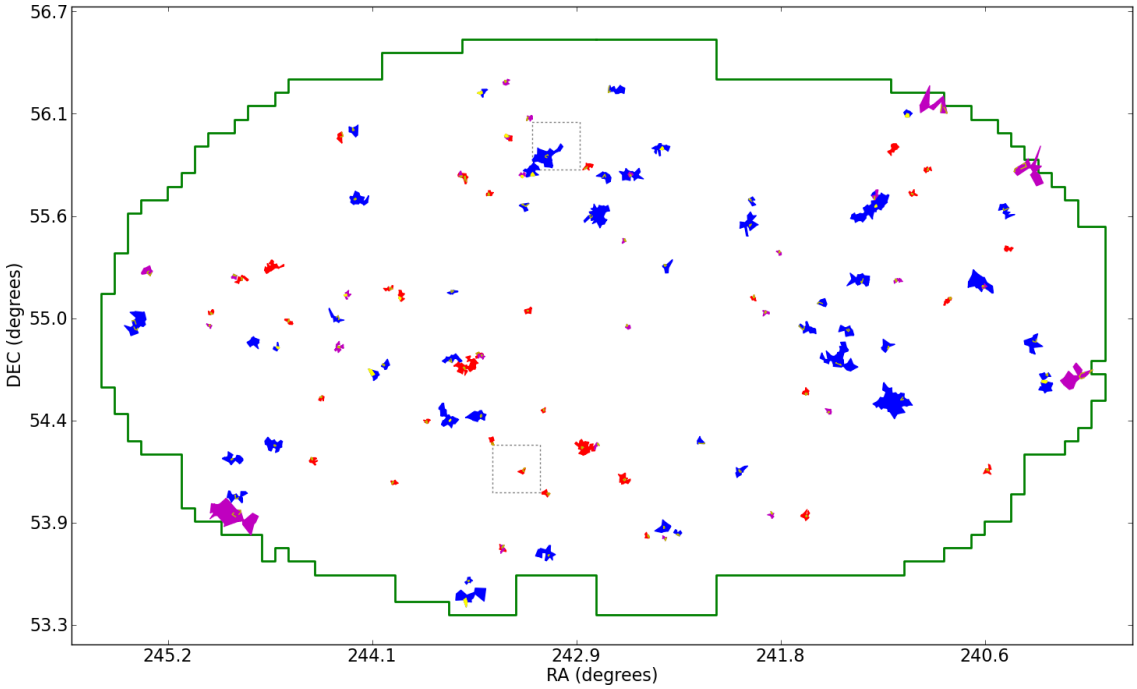

Figure 5.3: Clusters detected in the PS-1 MD-8 field. The coloured cells represent clusters identified in different colour pairs. Blue cells correspond to clusters detected in $\{g\text{-}r, r\text{-}i\}$ filter pairs, red cells in $\{r\text{-}i, i\text{-}z\}$ and magenta in $\{i\text{-}y\}$. Yellow cells belong to the BCG of each cluster. The green line delineates the boundary of the data as detected by DARB, and dotted grey squares denote the position of Abell clusters.

5.3.1 Hectospec spectroscopic followup

With access to PS-1 collaborator facilities, we were able to acquire spectra from the Hectospec instrument at the Magnum Mirror Telescope (MMT). Hectospec can sample a circular field of view 1° in diameter with 300 spectroscopic fibres. Some 30-50 fibres are left for sky calibrations, permitting approximately 250 simultaneous observations. For hour-long integrations, redshifts are measured down to a limiting magnitude of $r < 21.5$, although systems with emission lines may be detected faintwards of this limit. With a 270 line/mm grating, the 1.5"fibres can sample the 3650-9200Å spectral range with a resolution of about 5Å. At longer wavelengths, second-order light contamination becomes an issue, so the spectral shape long-ward of ~ 8500Å is subject to higher levels of noise.

At the time of writing, 20 BCGs have been targeted with Hectospec (R. Chornock, priv. communication), and redshifts computed with the automated pipeline. Of these

targets, two were stellar sources. In Figure 5.4 we show one of these mis-detections, with the *yellow* crosshair indicating the purported BCG and Hectospec target. Whilst visually it is apparent this object is stellar in nature, by coincidence it has a colour and magnitude that well fits the sequence of the group it contaminates. The *bottom* panel of Figure 5.4 shows colour-magnitude diagrams for *i*-*z* and *i*-*y* colours the group was identified in, with the yellow dot indicating the location of the star in colour-magnitude space. Although care was taken to remove stellar contaminants from the source catalogue, efforts to derive an optimal star-galaxy separation for Pan-STARRS data are ongoing. As such, this catalogue may unfortunately contain higher levels of stellar contamination than the final extended source catalogue.

In Table 5.2, we list the targets with pipeline-estimated redshifts and associated cluster redshift estimates. Because the initial sample containing these spectroscopic targets was selected by eye, their redshift distribution will be biased. However, in Figure 5.5 we show the accuracy of `ORCA` cluster redshift estimates. It can be seen that `ORCA` cluster redshifts systematically underestimate the spectroscopically calculated redshifts. The most distant systems also have larger redshift errors, indicating where photometric errors begin to dominate in the SDSS data `ORCA` redshifts originate from. With improved photometric redshifts based directly on deeper PS-1 data, these uncertainties should shrink. Figure 5.6 shows an example Hectospec spectrum for the BCG of cluster *MD08 J161315+56010.1*, the abscissa is the rest-frame wavelength in air. This BCG, (pictured in the *bottom* panel) has a redshift $\mathrm{z_{spec}} = 0.4892 \pm 0.0002$ that compares well with the calculated cluster redshift $\mathrm{z_{orca}} = 0.44 \pm 0.09$ estimated from photometric redshift estimates of five galaxies.

Due to a number of factors, it is likely the MD-8 galaxy catalogue is not yet ready for the identification of distant *y*-band cluster sequences. Firstly, Pan-STARRS has yet to reach the depth anticipated from the original specifications. Moreover, the simple `SExtractor` recipe for constructing the catalogue does not fully take into account the manner in which images are stacked by the Pan-STARRS pipeline. Finally, because sources are *i*-band selected, extremely red galaxies visible in the *y*-band lacking an *i*-band flux will not be selected. These factors suggest the current input catalogue is not optimal for the identification of high-redshift galaxies. Development work on the Pan-STARRS extended source photometry is ongoing. It is hoped once this is complete (and

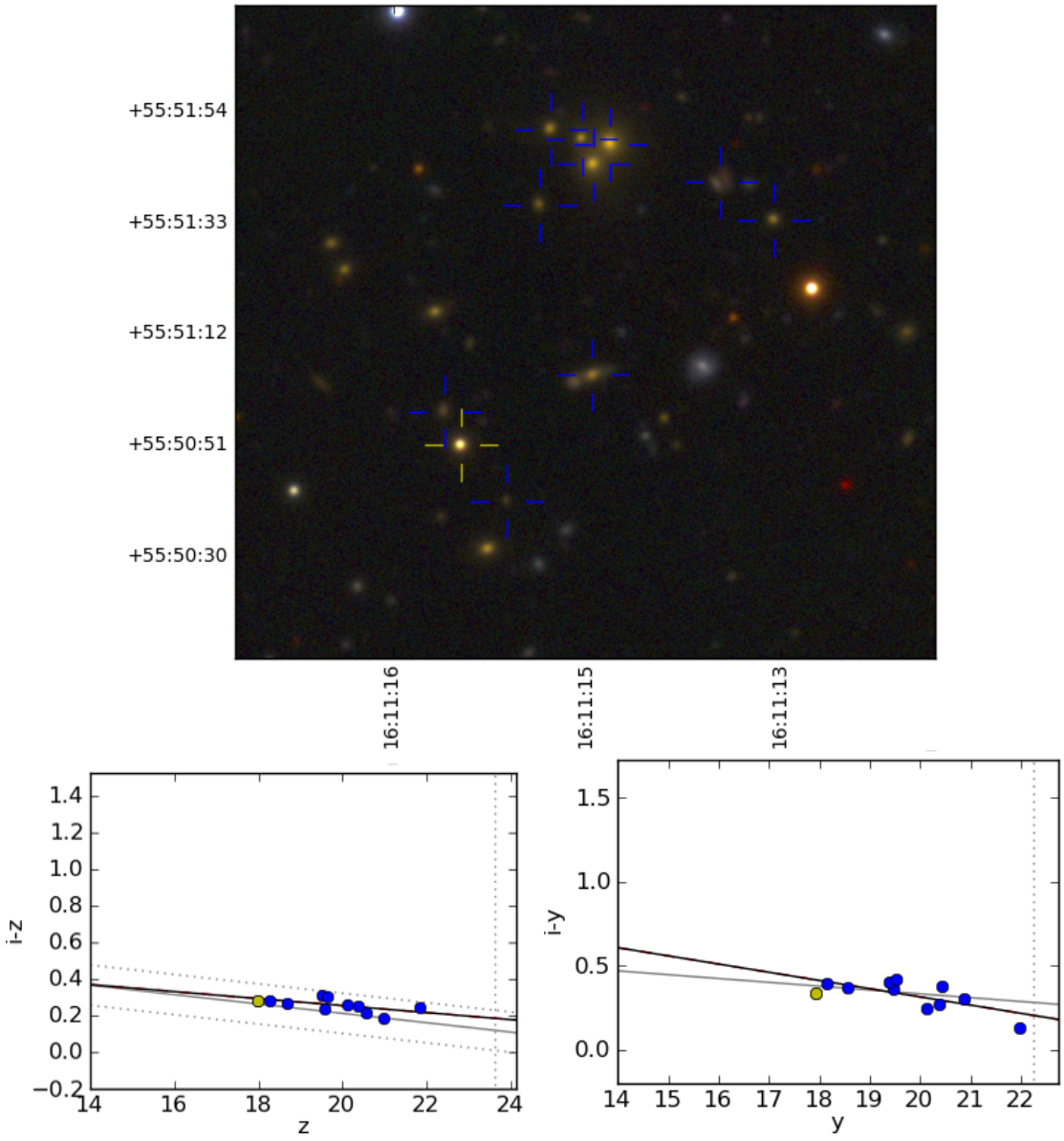

Figure 5.4: (*top*) *gri*-composite imaging of ORCA cluster *MD08 J161128+55510.4*. Cluster members are highlighted with blue crosshairs. The yellow crosshair indicates the BCG of the cluster, which in this case is a contaminating star. (*bottom*) The two red sequences used to select this cluster. The projected proximity of the stellar source to genuine cluster members, in addition to its observed colour and magnitude (shown as the yellow dots in the CMRs) conspire to establish this object as the cluster BCG. Vertical dotted lines are magnitude limits of the survey, horizontal grey dotted lines in the *i*-*z* CMR indicate the photometric selection filter the cluster was selected in. The black line in both plots show the best fit to the cluster sequence.

Name	z_{spec}	z_{orca}
MD08 J161247+56070.6	0.497	0.376 ± 0.158
MD08 J161315+56010.1	0.489	0.445 ± 0.088
MD08 J161257+55480.4	0.106	0.069 ± 0.014
MD08 J161128+55510.4	**0.000**	$\mathbf{0.236 \pm 0.126}$
MD08 J161105+55470.7	0.250	0.244 ± 0.040
MD08 J161031+55480.5	0.249	0.276 ± 0.072
MD08 J161115+55340.8	0.217	0.208 ± 0.049
MD08 J160951+55570.5	0.140	0.133 ± 0.070
MD08 J161226+55540.6	0.266	0.245 ± 0.071
MD08 J160748+55390.7	0.251	0.258 ± 0.046
MD08 J160502+55410.3	**0.000**	$\mathbf{0.542 \pm}$ ——*
MD08 J160412+55420.1	0.312	0.290 ± 0.045
MD08 J160506+55370.3	0.251	0.250 ± 0.048
MD08 J160522+55130.1	0.248	0.232 ± 0.058
MD08 J160612+55050.1	0.260	0.221 ± 0.029
MD08 J160445+54500.7	0.259	0.232 ± 0.044
MD08 J160129+54510.4	0.299	0.266 ± 0.048
MD08 J160431+55130.0	0.393	0.336 ± 0.115
MD08 J160539+54560.3	0.247	0.225 ± 0.021

Table 5.2: MD-8 ORCA clusters with Hectospec-measured redshifts (z_{spec}), along with the cluster redshift estimate (z_{orca}) derived from SDSS photometric redshifts. The contaminating stellar sources are emboldened in this table.

* only one galaxy with a redshift estimate for this cluster

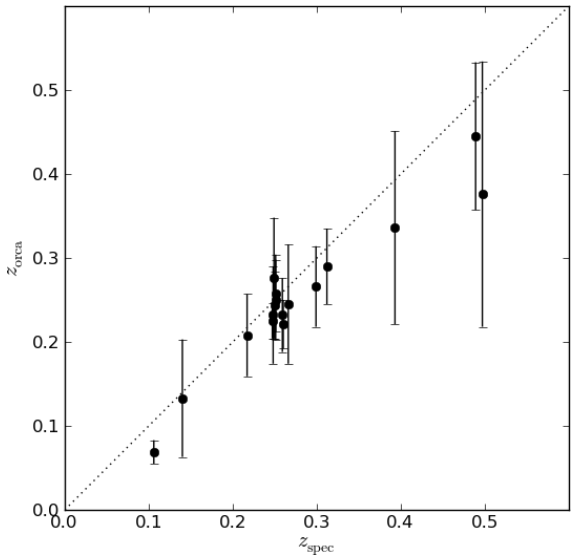

Figure 5.5: The accuracy of ORCA cluster redshift estimates generated from the photometric redshift data of the (shallower) SDSS, compared to the 18 galactic Hectospec targets recorded in Table 5.2.

by which time this field will have been re-visited), catalogues will be generated with more understanding of the photometric uncertainties, source selection and an improved star-galaxy separation. Until then, there are still projects using these cluster data that we hope will promote early extragalactic Pan-STARRS science. For example, a study of the Hectospec BCG spectra may reveal evidence of star formation in cluster cores. Additionally, the redshifts of these galaxies may be used to improve photometric redshift estimates.

5.4 Pan-STARRS Small Area Survey (SAS)

As part of the Pan-STARRS-1 commissioning process $\sim 65\text{deg}^2$ of imaging was released featuring *grizy* exposure times equivalent to those expected after three years of the 3π survey. Unfortunately poor observing conditions resulted in data of lower quality than expected. We stress that our findings in this section are therefore not a true representation of final PS-1 data. Extended source photometry was processed at Durham using SExtractor (P. Draper, priv. communication), this time with independent detections in each band. In this study, we compiled a band-merged source catalogue by matching, to within 1 arcsecond, *gizy* $3''$ aperture magnitude detections to *r*-band selected sources.

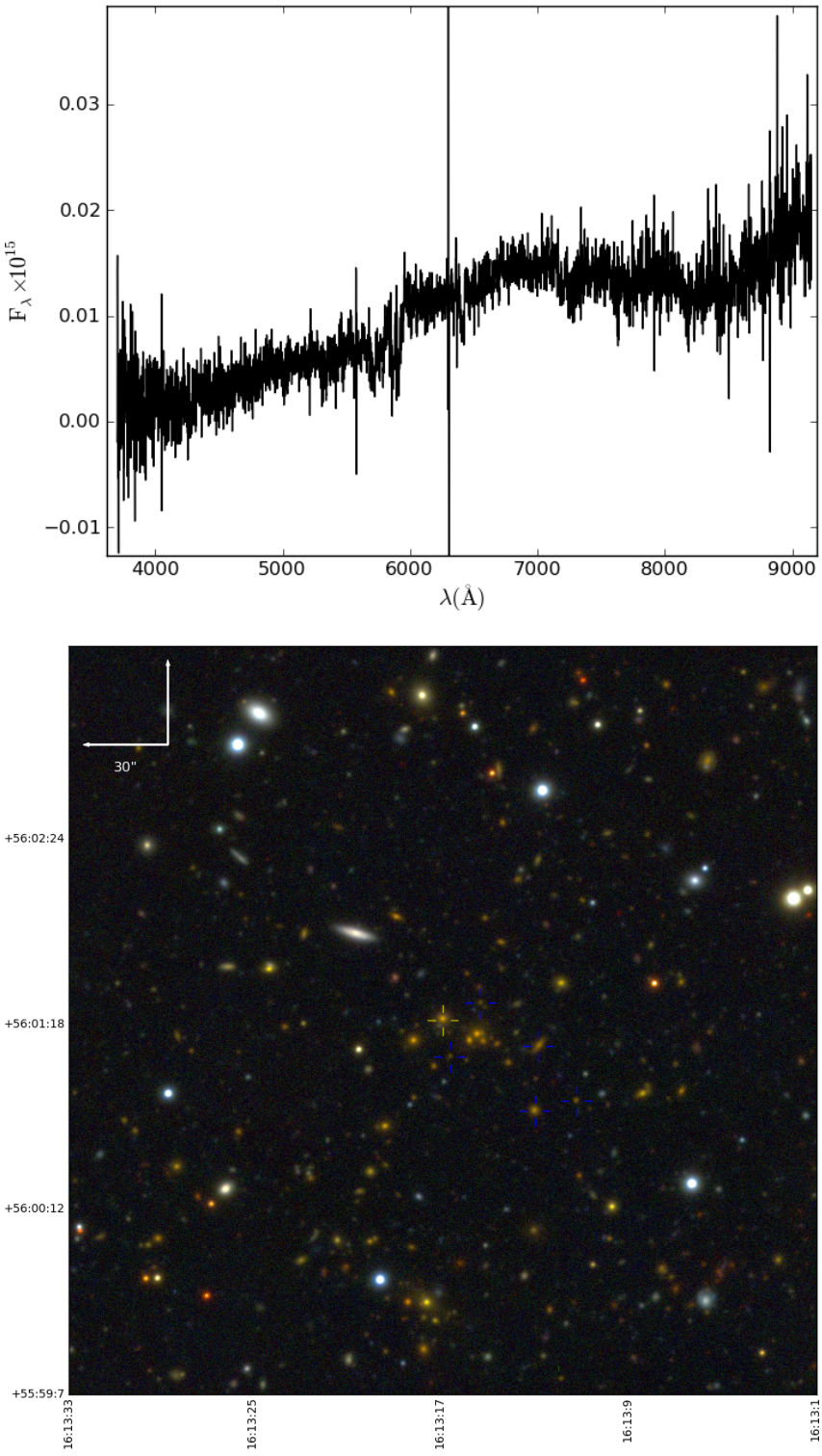

Figure 5.6: The Hectospec BCG spectrum (*top*) and Pan-STARRS MDS *irg* imaging (*bottom*) for cluster *MD08 J161315+56010.1*. Cluster members are identified with blue crosshairs. The BCG (yellow crosshair), has a spectroscopically measured redshift of $z = 0.489$.

With galaxy samples probing large depths, it is inevitable some galaxies will be neither measured nor detectable in all bands. For example, a very distant galaxy detected in the *y* band may not have any associated *g*-band flux (a "drop-out"); it is important to include both these and their bluer analogues. In this catalogue, some magnitudes may consequently not have any value assigned to them. For a forced-photometry approach akin to that used in the previous section, aperture magnitudes will be measured even where no source exists in that band. Typical catalogue convention is to set these unobserved magnitudes to a nonsensical figure - often ± 99. For ORCA to work on photometry with both independently measured bands and drop-outs, we relax the galaxy-colour rejection constraint. A galaxy with measured flux in two consecutive bands, producing a physically realistic colour, is included in the source catalogue.

The *y*-band magnitudes were calibrated to the SDSS z-band, and as such required an offset to recover the true *y*-band magnitudes. Because the *z*-*y* colour is sensitive to distant cluster sequences beyond the depth of this survey ($z \gtrsim 0.9$), dropping this band resulted in no loss of detection capability and furthermore allowed direct comparison to the SDSS *griz* photometry. Because the survey depth is comparable to Sloan photometry covering the same region, we cross-matched PS-1 sources with SDSS redshift data; Pan-STARRS galaxies inherited the redshifts of any SDSS match closer than $1''$.

A total of 649,668 galaxies were merged into a source catalogue. Magnitude limits, set by the procedure detailed in §3.3.7, were calculated as (g, r, i, z)=(22.7,22.6,21.7,21.0). We scan the data with DARB to determine the boundary edges and holes (arising from incomplete data in some bands), and determine a survey area of 61.8deg^2 . We set the algorithm to search through the same colour range explored in the SDSS Stripe 82 catalogue, using the identical photometric selection parameters (β and σ, defined in Table 3.2) for the {*g*-*r*, *r*-*i*} and {*r*-*i*, *i*-*z*} filter combinations. This scan allocated 1,233 galaxies into 153 clusters. In Figure 5.7, we plot the position of these cluster detections, along with the survey boundary (*green*) and the six Abell clusters (*grey* dotted squares) identified in this field. We provide details of the cluster properties in §C.4 of Appendix C.

Using DARB, we apply the PS-1 data boundaries (including survey holes) to SDSS DR7 data, creating an identical and equivalent footprint with which to directly compare against PS-1 detections. Because these data consist of single-epoch SDSS observations, we cannot use deeper Stripe 82 magnitude limits in this study. Instead, we determine

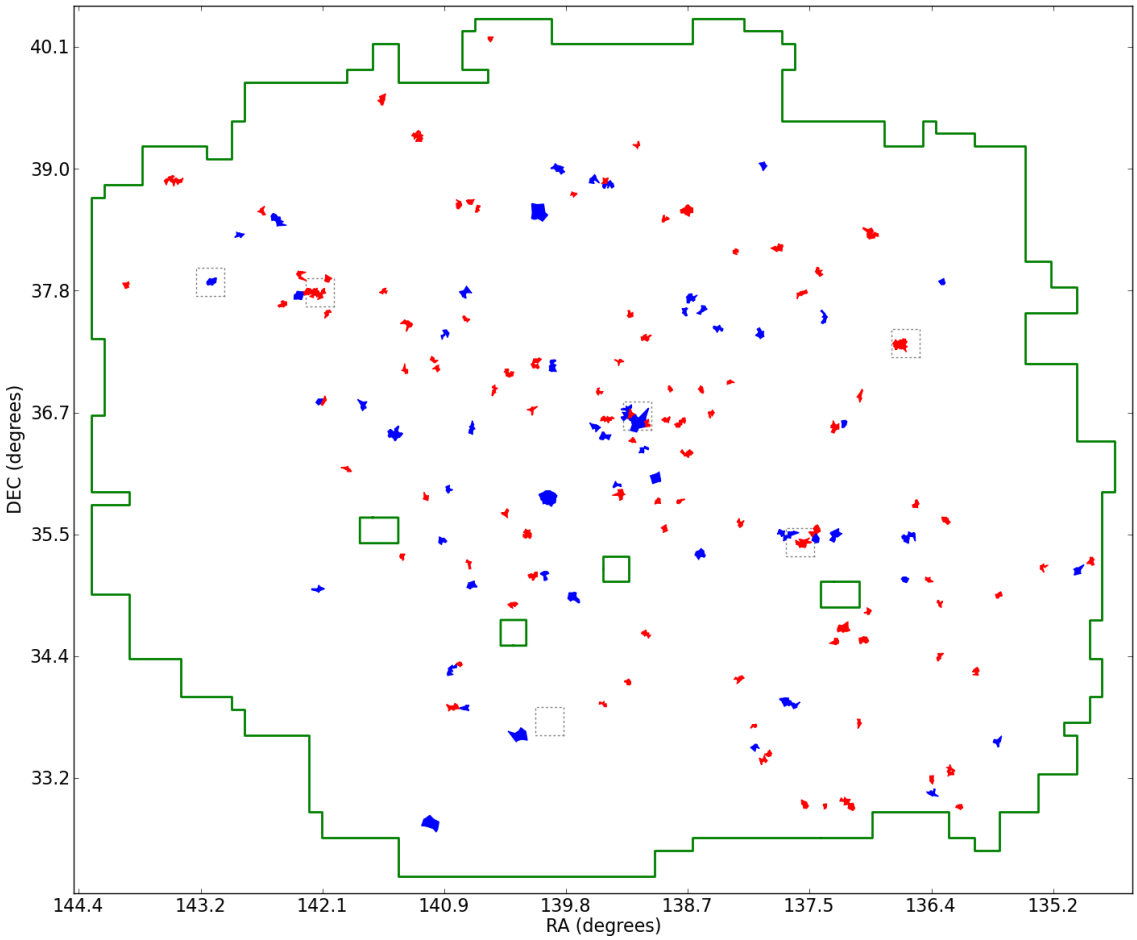

Figure 5.7: Clusters detected in the PS-1 SAS field. The coloured cells represent clusters identified in different colour pairs. Blue cells correspond to clusters detected in {*g*-*r*, *r*-*i*}, red cells in {*r*-*i*, *i*-*z*}. The green lines describe the boundaries of the data, including four "holes" detected by DARB . Dotted grey squares denote the position of Abell clusters.

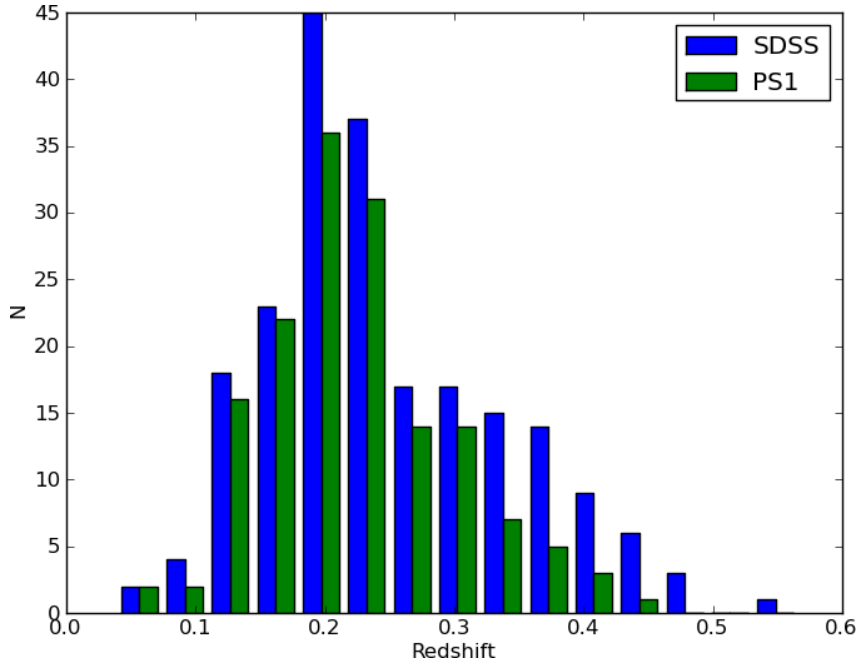

Figure 5.8: The redshift distribution of PS-1 (green) and SDSS DR7 (blue) galaxy clusters in the Pan-STARRS SAS survey region.

the magnitude error distribution in a manner consistent with the PS-1 (and Stripe 82) analysis, arriving at limits of (g, r, i, z) = (22.6,22.2,21.7,20.1) (see section §5.6 later for more details). From this catalogue, we identify 211 galaxy clusters in the SAS footprint. In Figure 5.8 we compare the redshift distributions of the two cluster catalogues. SDSS data, in addition to identifying more galaxy clusters, also has a slightly higher median cluster redshift (0.22 vs. 0.21). As Figure 5.8 shows, this can be attributed to the more extended high-redshift tail. Both this tail and the higher cluster yield appears to suggest the SDSS data is either deeper (despite brighter flux limits applied to Sloan data), of higher photometric quality, or both. In the *top* panel of Figure 5.9 we again show the clusters detected with Pan-STARRS, but with SDSS detections represented by *grey* circles (radii denoting the peak centre-member separation). The Sloan data extends into the south-eastern part of the survey, and in general appears more evenly distributed. It is likely that the Pan-STARRS data is not of a uniform depth across the field. In the *bottom* panel of Figure 5.9 we show the distribution of galaxies within a narrow 0.2 magnitude slice near the faint i-band limit of the PS-1 data, below a magnitude error motivated from our choice of photometric filter width. The points, colour-coded by i-band error are not

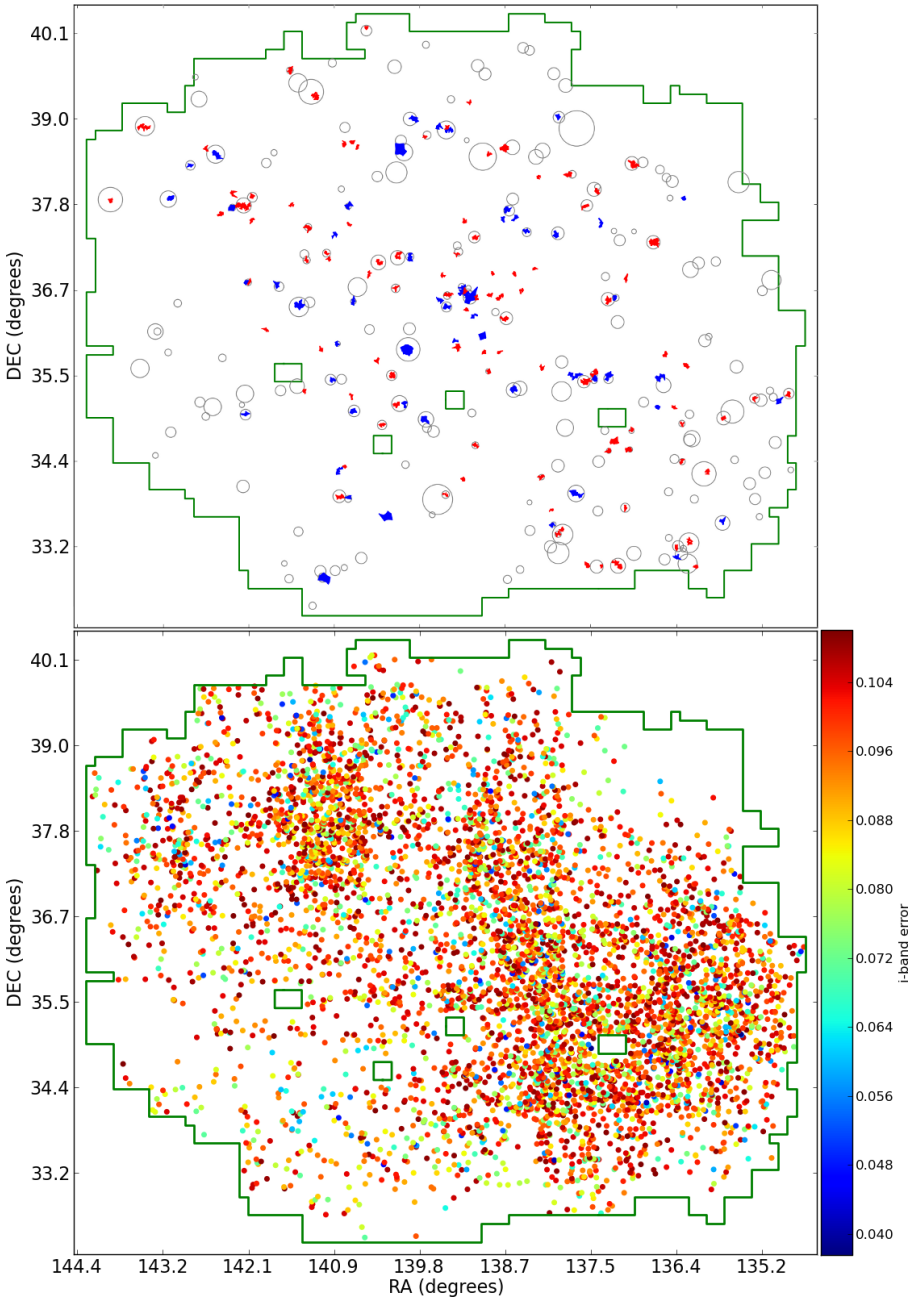

Figure 5.9: (*top*) As for Figure 5.7, but with the positions of SDSS ORCA detections marked in this field as grey circles. (*bottom*) The distribution of Pan-STARRS galaxies in a narrow magnitude range close to the i-band limit of the survey. Points are colour-coded according to their magnitude error, with red indicating errors close to the limit Δm=0.117. This plot suggests the galaxies are not uniformly distributed across the survey.

uniformly spread and appear to correlate with the density of cluster detections in the upper panel.

At the cluster densities derived from this study, we recover only 75% of the number anticipated by the Pan-STARRS Design Reference Mission (DRM; approximately 10^5 over the whole survey). Whilst this preliminary investigation appears pessimistic, as discussed at the start of this section it was later determined imaging data was taken under poor photometric conditions unrepresentative of the anticipated 3-year quality. Moreover, the source catalogue constructed for this investigation used fixed aperture photometry with magnitude errors perhaps not truly estimating the actual uncertainty in detected flux. With improved imaging and extended source photometry from the Pan-STARRS Image Processing Pipeline (IPP), future studies along this theme may prove useful for benchmarking Pan-STARRS data against the SDSS.

5.5 Mass calibration in Pan-STARRS data

Two parameters key to the scientific analysis of clusters are their redshifts, and an estimate of their mass. Because one cannot directly measure the cluster mass, alternative observables must be used as proxies that correlate with mass. With access only to photometric optical data, the most direct mass estimator is the total galaxy luminosity[1], and its relation to the stellar mass within the cluster. Because in turn total stellar mass correlates with the mass of the halo hosting the cluster (see the *bottom-right* panel of Figure 5.22 shown later in §5.7), the combined galaxy luminosity can act as a tracer of cluster mass. In this section we will estimate cluster masses from PS-1 MD-5 data, based on calculations of the y-band cluster luminosity.

The Pan-STARRS y filter (with central wavelength $\lambda_c \simeq 9700$Å) shares a common feature with near-IR bands: it can trace stellar mass with only minimal contaminating contribution from the star-forming population and dust, allowing for accurate determinations of stellar mass in galactic systems, even at high redshift (see e.g. Kodama and Bower, 2003). In Figure 5.10 we plot the observed-frame cluster M_*-L_y relation (colour-coded according to cluster redshift) using Bower et al. (2006) semi-analytic galaxy data

[1]One may alternatively measure the weakly-lensed shear of background galaxy shapes, but this relies on very deep data taken in sub-arcsecond observing conditions.

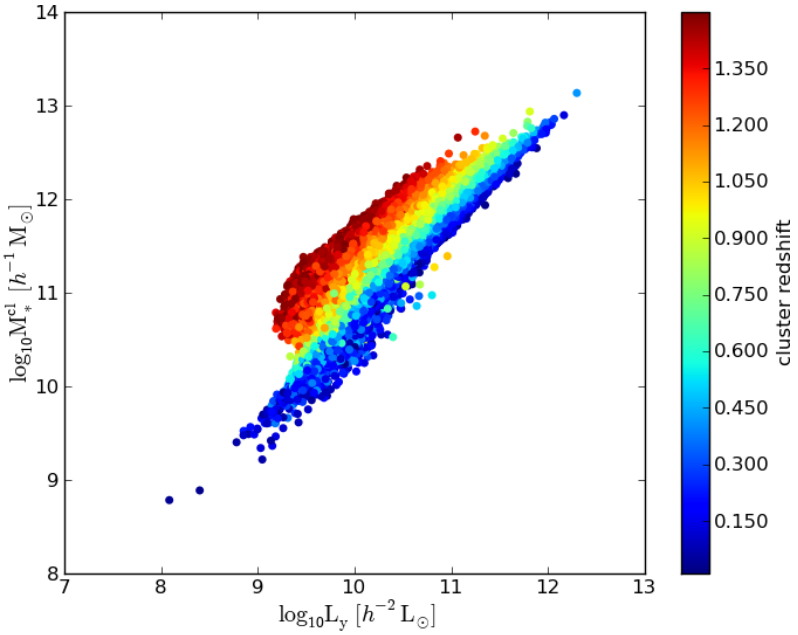

Figure 5.10: The relationship between y-band luminosity and stellar mass for clusters in a lightcone simulation. Each point, corresponding to a cluster, is colour-coded according to the redshift of the dark matter halo it resides in.

from a 100 square degree lightcone (A. Merson, priv. communication).

After determining the galaxy membership of the haloes from the lightcone dark matter data, for each filter we calculate the total (bulge and disk) stellar mass and y-band luminosity. Although this simple analysis does not include survey flux limits or photometric errors, comparing such correlations against observed data (such as stellar masses from spectroscopy) is a powerful test of galaxy formation model.

We demonstrate how y-band mass estimates can be made using a cluster catalogue of ORCA detections from the Pan-STARRS MD-5 field (commonly known as the "Lockman Hole", covering 11.6deg^2 centred at α, δ=161.917°, 58.083°). At the time of writing, this field benefited from deeper photometry than the MD-8 field presented in §5.3. In brief, ORCA identified 133 clusters and 1342 cluster galaxies from $grizy$ photometry down to limits of 25.15, 24.5, 25.2, 24.5 and 23.40 respectively.

A series of linear fits to the data presented in Figure 5.10 were made in redshift bins of width $\mathrm{dz} = 0.05$ out to $\mathrm{z} = 0.7$. Within each redshift bin, the level of luminosity-dependent scatter in estimated M_* indicates the level of mass uncertainty at this redshift.

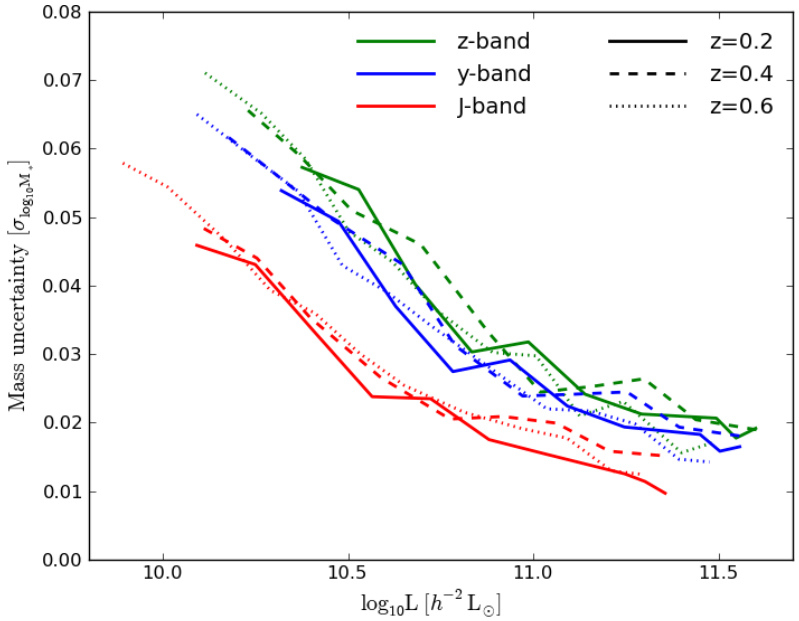

Figure 5.11: Uncertainties in cluster stellar mass from total galaxy luminosities in the z, y and J bands at three redshifts.

For three redshift bins, we compare this scatter in stellar mass to the equivalent set of correlations in the SDSS z and UKIDSS J bands. Because the z-filter ($\lambda_c \simeq 9100$Å) probes a slightly bluer region of each galaxy SED, we expect more sensitivity to star formation and dust, increasing the scatter in the M_*-L_z relation. Conversely the J filter ($\lambda_c \simeq 12{,}600$Å) measures a much redder portion of the galaxy SED, thus reducing scatter in the mass estimate. In Figure 5.11, we compare the scatter in stellar mass $\sigma_{\log_{10}M_*}$ for the three bands at three redshifts. As anticipated, the y-band offers an improvement over the z-band estimates, but exhibits a higher scatter compared to the near-IR J-band. At high cluster luminosity, differences between the optical-band mass estimates are marginal; at lower masses, there is a clear advantage in adopting the y-band over the z-band.

To estimate the total y-band cluster luminosity in Pan-STARRS data, in each cluster we simply sum the luminosities of all galaxy members detected by `ORCA` . However, to improve our estimate, below we consider four additional sources of uncertainty:

1. *Unobserved luminosity from galaxies outside the survey flux limits*: To correct for galaxies unobserved due to the survey flux limit, we adopt the approach laid out in §2.4.1. Assuming the luminosity function of cluster galaxies follows a Schechter (1976) func-

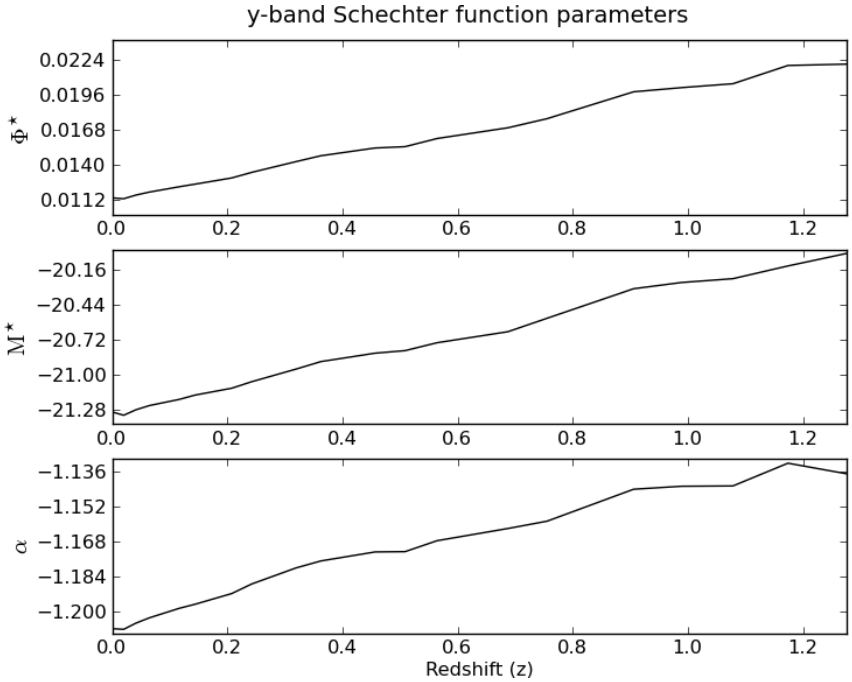

Figure 5.12: Redshift evolution of the observed-frame y-band Schechter function parameters from Bower et al. (2006) semi-analytic model galaxies in Millennium Simulation snapshots to z$\leq$1.28.

tion, we re-scale each observed-frame galaxy luminosity $L_{i,y}$ to $L_{i,y}^{\rm corr}$ in a similar manner to Equation 2.8:

$$L_{i,y}^{\rm corr} = \frac{L_{i,y}}{\Gamma[\alpha(\mathrm{z})+2, L_{max}/L^\star(\mathrm{z})] - \Gamma[\alpha(\mathrm{z})+2, L_{min}/L^\star(\mathrm{z})]}, \tag{5.1}$$

where Γ is the incomplete Gamma function, $\alpha(\mathrm{z})$ and $L^\star(\mathrm{z})$ are the redshift-dependent y-band Schechter function parameters, L_{max} and L_{min} correspond respectively to the bright (13.3, set from the brightest galaxy in the source catalogue) and faint (23.40) y-band flux limits. The luminosity function redshift evolution was determined from Schechter fits to the luminosities of Bower et al. (2006) semi-analytic model galaxies, populated into 20 snapshots between `snapnum` 38 (z$\simeq$1.28) and 63 (z=0) of the Millennium Simulation. Figure 5.12 shows the redshift evolution of the luminosity function parameters, where one may convert between $L^\star(\mathrm{z})$ and $M^\star(\mathrm{z})$ with $\mathrm{M}_\odot$=4.025[2].

2. `ORCA`-*detected galaxies without measured y-band fluxes*: This particular correction only arises when independent detections from each band are positionally matched (as with

[2]In the absence of a measured y-band $\mathrm{M}_\odot$, this value was estimated from a linear interpolation between the z and J bands.

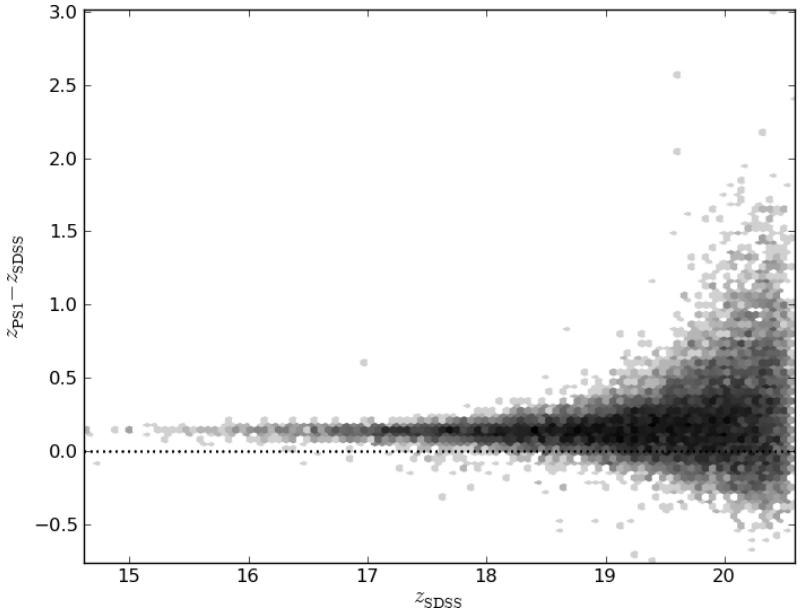

Figure 5.13: The zeropoint error in PS-1 z-band photometry when compared to SDSS. The intensities are logarithmically-spaced, and show the number of galaxies at each point.

the SAS survey presented in §5.4). To provide statistical corrections for cluster galaxies without y-band fluxes, we randomly assign each galaxy a luminosity weighted by the Schechter function corresponding to the cluster redshift. This luminosity is subsequently corrected for survey flux limits as described above. In our MD-5 cluster sample, 33 galaxies (2.5%) did not have y-band magnitudes, and were consequently assigned luminosities.

3. *Stellar mass from genuine cluster galaxies that* ORCA *missed*: Because inevitably ORCA will not identify all members in a cluster, we must make a correction for the stellar mass incompleteness of each detected system. Our analysis of the mock MDS data in §3.7 included a measurement of the stellar mass accuracy A_* (see Figure 3.25) - the fraction of recovered stellar mass in the mock clusters. We determine (by averaging over the mass bins in that Figure) the redshift-dependent stellar mass accuracy, and multiply our estimated cluster stellar mass by a factor $1/A_*(\mathrm{z})$.

4. *Errors in the y-band zeropoint*: Any errors in the y-band zeropoint will systematically bias estimates of galaxy luminosities. Zeropoints in the MD-5 catalogue studied here have been calibrated by the PS-1 Image Processing Pipeline (IPP). This procedure

involves fitting synthetic stellar spectra from 2MASS JHK (Cohen et al., 2003), USNO-B (Monet et al., 2003) and Tycho (Bessell, 2000) photometry, and deriving $grizy$ photometry with the SEDs. To determine how accurate this approach was, we match all MD-5 galaxies with z and y-band measurements to SDSS DR7 galaxies covering the same field. By plotting the magnitude difference against the SDSS aperture magnitude, in Figure 5.13 we compare the 3 arcsecond z-band aperture magnitudes of 15,469 matched galaxies between PS-1 and SDSS. This Figure shows PS-1 systematically underestimates the z-band galaxy flux by $\Delta m_z \simeq$0.1 magnitudes, in line with independent measurements of Δm_z=0.07 from SDSS and PS-1 3π survey point sources (P. Draper, priv. communication). The presence of a gradient in this offset indicates that a "colour term" must be introduced into equations converting between these two filters. Assuming the slightly redder y-band magnitudes are similarly underestimated in the IPP implementation, we adjust each y-band cluster galaxy magnitude by -0.1 mag to account for this zeropoint error.

With these corrections, we are now able to accurately estimate the cluster stellar mass for each cluster ORCA identified in the MD-5 field. We find cluster stellar masses spanning a range $10^{10.43} \leq \mathrm{M}_*^{\mathrm{cl}} h^{-1} M_\odot \leq 10^{12.50}$, with a median mass of $10^{11.49} h^{-1} M_\odot$. Only in one case (due to the lack of redshift data for a cluster redshift) were we unable to measure the cluster mass. An estimate of the halo mass via the model M_*-$\mathrm{M}_{\mathrm{halo}}$ relation depicted in Figure 5.22, though possible, is not presented here.

It is highly useful to compare these mass estimates with observables known to correlate with cluster mass. In Figure 5.14 we show two such correlations in the MD-5 (*red* points) cluster catalogue: $\mathrm{N}_{\mathrm{galaxies}}$ - the number of galaxies detected by ORCA, and $A_{\mathrm{gc[1,i]}}$ - the angular correlation function amplitude calculated out to θ_{80} for galaxies in the photometric selection filters the cluster was identified from (referred to in §4.4.2). We compare both of these mass-observable correlations to those calculated from the mock MDS survey used in Chapter 3 to evaluate the ORCA detector. Because in some cases the $A_{\mathrm{gc[1,i]}}$ estimator fails (for example due to Poisson error in N_T and N_B with small systems), some clusters may have nonsensical (i.e negative) richnesses. We have omitted these from Figure 5.14, but note the observed cluster catalogue has a higher $A_{\mathrm{gc[1,i]}}$-estimation failure rate (19%) than the mock (2%).

The mock and observed data show similar levels of scatter at low $\mathrm{N}_{\mathrm{galaxies}}$ and low

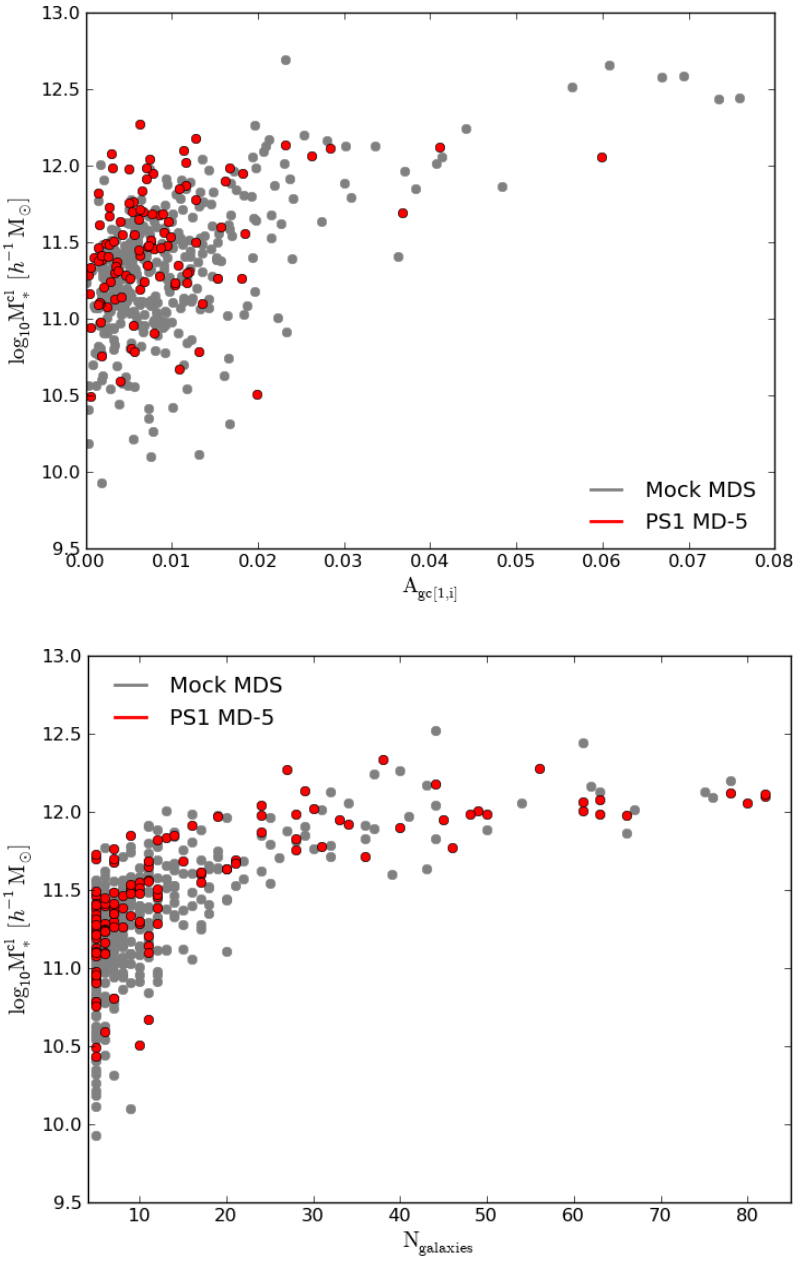

Figure 5.14: Correlations between cluster stellar mass and two measures of cluster richness. In the *top* panel, we compare cluster stellar mass to the angular correlation function amplitude $A_{\mathrm{gc[1,i]}}$ for the MD-5 (red) and mock MDS (grey) surveys. In the *bottom* panel, we compare the cluster stellar mass to the number of galaxies `ORCA` assigns to the cluster.

$A_{\rm gc[1,i]}$, suggesting mass estimates at the group scale remain challenging despite access to y-band data. The $\mathrm{N_{galaxies}} - \mathrm{M_*^{cl}}$ relation shows a stronger correlation and better agreement between model and observations, suggesting at $\mathrm{N_{galaxies}} \geq 10$ a simple linear fit could be used to accurately estimate (where no y-band measurements are available) the cluster stellar mass from `ORCA` membership. Although $A_{\rm gc[1,i]}$ does not correlate as strongly, future improvements to optical richness definitions may reduce the scatter for small systems and improve the estimation failure rate. With access to the PS-1 3π survey, future work can correlate y-band cluster masses to, for example, SZ decrements from Planck (Planck Collaboration et al., 2011) and X-ray temperatures from *Chandra* (Maughan et al., 2008) and *XMM-Newton* (e.g. Mehrtens et al., 2011). A table detailing the properties of each MD-5 cluster (including stellar mass estimates, with $A_{\rm gc[1,i]}$ replacing $B_{\rm gc}$) can be found in §C.5 of Appendix C.

5.6 Sloan Digital Sky Survey DR7

We now turn our attention to the SDSS DR7 photometric dataset as a testing ground for the detection of clusters in the next generation of wide-area deep surveys. At the time of writing, this survey represents the widest-area optically-selected galaxy catalogue available. The large SDSS contiguous region (CR), covering $\sim 7,500\mathrm{deg}^2$ represents $\sim 30\%$ of the area to be covered by the Pan-STARRS 3π survey. For the Sloan data, a drift-scan approach is used to image the sky whereby the camera takes 55 second exposures at the sidereal rate in a series of strips. In combination, these strips form stripes 2.5° wide in declination following great circles in the sky. The SDSS collaboration, through this approach, have compiled the large region forming the focus of our study in this section.

5.6.1 The data

The query used was the same as that of Stripe 82 (see Appendix B), but instead used the DR7 database. Although in the single epoch data this returns sources with very high levels of photometric uncertainty, we apply magnitude cuts to select the best subset whilst better characterising the noise characteristics of our sample. We take data from stripes[3] 9 to 39, corresponding to declinations between -4° and 74°, selecting all galaxies

[3]Stripe number is related to declination by $\delta = -25 + 2.5s$, such that (s)tripe 10 is equatorial

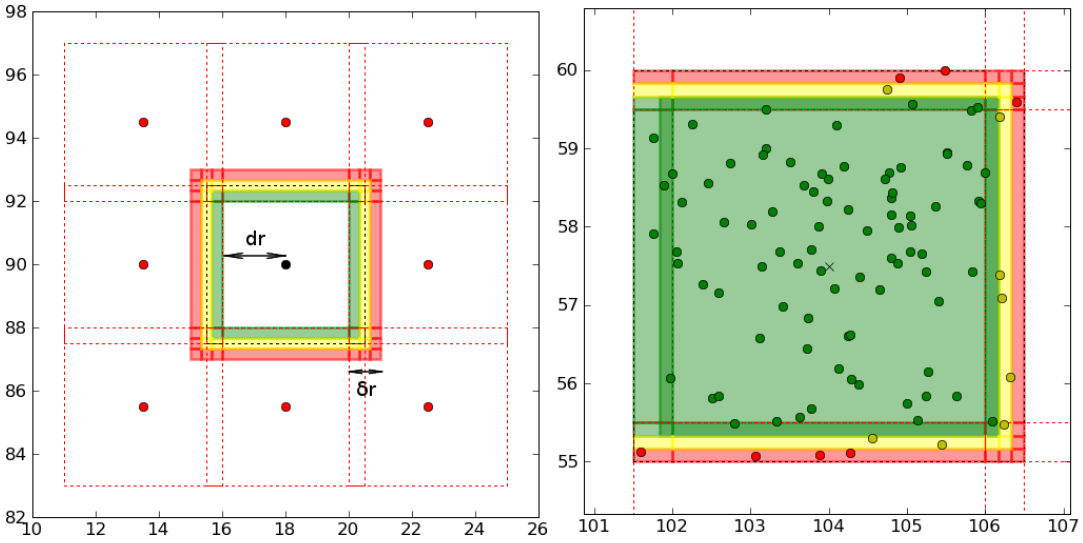

Figure 5.15: (*left*) The partitioning scheme applied to SDSS DR7 data. In general, each cell is bordered by eight other cells. These cells partially overlap with the central cell in both RA and dec. The coloured strips indicate this overlap zone (δr in width). (*right*) An example cell at the edge of a data grid (such as cell 166 in Figure 5.16). As with the Stripe 82 scheme, three coloured regions indicate the status of a cluster detected in these regions: always kept (green), kept if no nearby or "better" candidates exist (yellow), always rejected (red).

between $105^\circ \leq \alpha \leq 265^\circ$ in a series of blocks in right-ascension. This safeguarded against submission timeouts and failed jobs arising from exceeding the allocated quota space for SQL queries. Each block overlapped slightly with a neighbour to ensure no galaxies were missed at the boundaries. Once all RA blocks were merged and duplicate galaxies removed, the source catalogue contained 110,050,628 galaxies. Before applying ORCA to this dataset, we make a series of improvements to the algorithm in light of investigations subsequent to our Stripe 82 analysis - these are detailed below.

5.6.2 Partitioning the data

In §5.4 we performed both PS-1 and SDSS detections over $\sim 65\text{deg}^2$. While these runs completed in a timely manner, it is unfeasible to scan a field two orders of magnitude larger in one chunk, and the depth across the whole footprint is not uniform. Our alternative approach is to extend the technique applied to Stripe 82 data in §4.3.2 by dividing

the survey into cells.

The *left* panel of Figure 5.15 shows an example cell surrounded by 8 neighbours. To ensure no clusters are divided by the cell boundaries, each cell is partially overlapped by its neighbours. The half-length of the cell base out to the boundary region is denoted by dr in the Figure; the overlap boundary has width δr. We must therefore define a scheme to merge clusters detected in each of these cells, ensuring a cluster identified in a shared area is counted only once. We apply a similar approach to that of the neighbouring sectors in the Stripe 82 analysis. This is depicted in the *left* panel of Figure 5.15 - the coloured borders highlight the areas shared between the central cell and its neighbours, split into three zones. Concentrating on detections specifically from the central cell, clusters found in the *green* zone are automatically added to the final catalogue. Conversely, clusters found in the *red* zone are never added. The *yellow* zone acts as a transitional buffer between the two cells.

The *right* panel of Figure 5.15 shows this scheme in progress, this time for a test cell at the edge of the grid containing randomly distributed points colour coded according to their status. We apply a grid of these cells to the SDSS data, setting the cell side length to be 9.5° (including a $45'$ overlap buffer) meaning each cell has an area 90.25deg^2 (dr, δr = (4°, 0.75° in Figure 5.15). Based on the footprint of the data, this results in the generation of 152 cells seen in Figure 5.16. The gradient observed in the data shown in the *top* panel of this Figure arises from the conformal projection used in the gridding - this is discussed further in §5.6.5.

5.6.3 Merging of multiple cluster detections

Following the detection of clusters in any of these cells, their overlaps must be checked with the part of the algorithm identifying multiple detections of the same cluster. In an effort to reduce the accidental inclusion of identical clusters, we revise the thresholds originally set in Table 3.1 to those in the Table 5.3 (updated values have been emboldened). These new thresholds relaxed the criteria for merging two clusters, and were arrived at by careful analysis of sample clusters in cell 92. By comparing detections to visual *irg*-composite images, clusters perceived to have been either fragmented or duplicated were studied under relaxed limits to establish the thresholds at which they

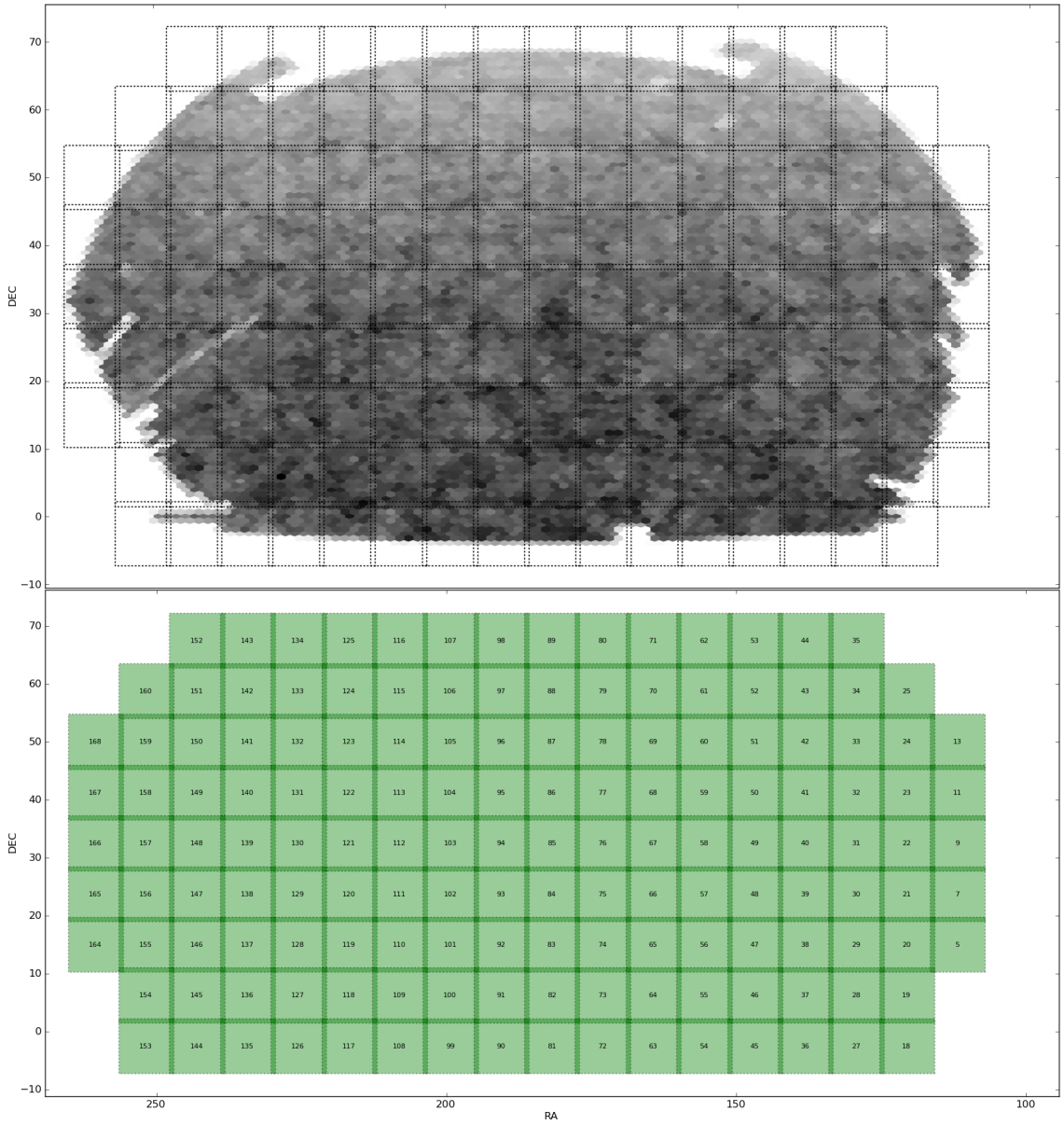

Figure 5.16: (*top*) The SDSS CR footprint, with grey-scales indicating the projected density of galaxies within the flux limits defined in §5.6.4. Superimposed on this plot is the grid scheme applied to partition the survey (dotted lines). (*bottom*) The same grid, with the cell ID numbers shown. Cell 92 (α =190°, δ=15°) was used as a test cell to improve the merging of multiply-detected clusters.

#	Constraint
1	$(cg_{1,2}$ OR $cg_{2,1}) \geq \mathbf{0.2}$
2	$(pe_{1,2}$ OR $pe_{2,1}) > 0$ AND ΔS_1
3	BCG_{id} AND ΔS_2
4	$(pe_{1,2}$ OR $pe_{2,1}) \geq \mathbf{0.2}$ AND ΔS_3
5	$(pe_{1,2}$ OR $pe_{2,1}) \geq \mathbf{0.2}$ AND ΔS_4

Table 5.3: New thresholds (emboldened) set following investigation of merging and fragmentation of ORCA clusters. These conditions are used to consider whether two clusters are multiple detections of the same system. If any one of these conditions are satisfied, the algorithm picks the "best" cluster of the two.

merged.

Previously, where a choice was to be made between detections of the same cluster, the BCG was simply the brightest galaxy in the "best" candidate. In some cases this particular cluster may feature a bright peripheral galaxy classified as the BCG, whilst alternative detections had instead designated a more central member, assumed closer to the minimum of the cluster potential. In the more relaxed BCG allocation scheme that follows, we make use of *associate cluster members* - those galaxies not in the final selected cluster, but nevertheless assigned to one or more of the candidate clusters.

We select the brightest cluster member (associate or otherwise) within $0.5\mathrm{r}_{\mathrm{max}}$ of the "best" cluster's estimated centre. This radial constraint ensures the inclusion of a genuine but peripheral cluster member does not prevail over central systems more traditionally associated as cluster BCGs. This relaxation in BCG definition is also motivated by the observation that central galaxies are often bluer than the red sequence formed by their companions. A visual example can be seen in Figure 3.21 (chapter 3), possibly arising from low-level star formation as gas is processed in the cluster core. The new BCG assignment procedure in effect widens the acceptable colour range a BCG may occupy[4] by considering associate members detected in filters flanking the cluster sequence.

[4]Although only for those clusters detected in more than one photometric filter - approximately 70% of clusters in final catalogues.

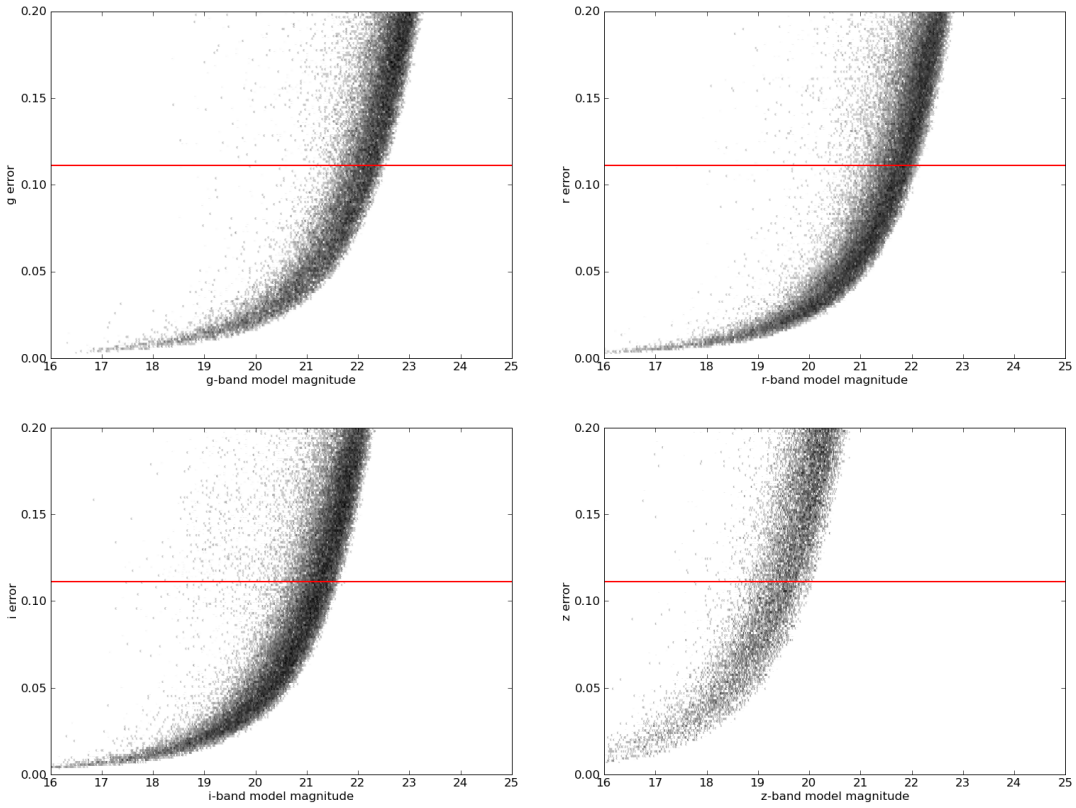

Figure 5.17: The model magnitude errors as a function of magnitude for the *griz* SDSS bands. These are used to formulate magnitude limits before the cluster detector is applied to the data. The horizontal red line indicates the threshold above which we exclude, by setting a suitable magnitude limit, galaxies with significant photometric uncertainties.

Colour	Slope (β)	Width (σ)	Range	1^{st}-stage filters
g-r	variable	0.152	$0.47 - 2.00$	39
r-i	variable	0.152	$0.00 - 1.22$	38
i-z	variable	0.152	$-0.10 - 1.10$	31

Table 5.4: Photometric selection filter parameters applied to the DR7 data.

5.6.4 Photometric filters

We now turn our attention to the set of photometric filters that will be applied to each cell. As with detections in Stripe 82, we will employ the *g-r*, *r-i* and *i-z* band-pairs to identify sequences. We keep the same selection filter widths and colour ranges, summarised in Table 5.4. Magnitude limits are applied to the input catalogue in order to prevent the inclusion of photometrically noisy data. We apply the same approach discussed in §3.3.7 and apply cuts of (g, r, i, z) = (22.6,22.2,21.7,20.1) to extinction-corrected galaxies. The magnitude error distribution for these four bands can be seen in Figure 5.17. These cuts reduced the source catalogue to 23,829,607 galaxies.

We make use of our deeper detections from chapter 4's Stripe 82 data to produce filter slopes (β) that vary with normalisation ($c_{\rm m20}$) when searching through the colour-magnitude space. In general, the filter width should remain broad enough to encapsulate the entire sequence of a speculative cluster. However, adjusting the slope to better represent the predicted cluster sequence gradient may, in reducing the selection of background galaxies, increase the contrast of the cluster signal. To determine $\beta(c_{\rm m20})$ for all three colours, we fit splines to the $c_{\rm m20}$-β relation using the following procedure. Sequence slopes and colour normalisations are calculated for every Stripe 82 cluster in the colour they were detected in. In each colour, distributions sampling these data were generated by calculating the median sequence slope of clusters in a set of $c_{\rm m20}$ bins. A spline fitted[5] to these points quantified the sequence slope evolution with normalisation. To improve spline fits, each $c_{\rm m20}$ bin was assigned a weight $1/\sigma$, with σ calculated as the scatter of cluster sequence slopes within the bin. To prevent bins with small samples influencing the fit, we reject those containing fewer than five clusters. Figure 5.18 shows

[5]We use the B-spline fitting routine from the FORTRAN `FITPACK` library (Dierckx, 1993).

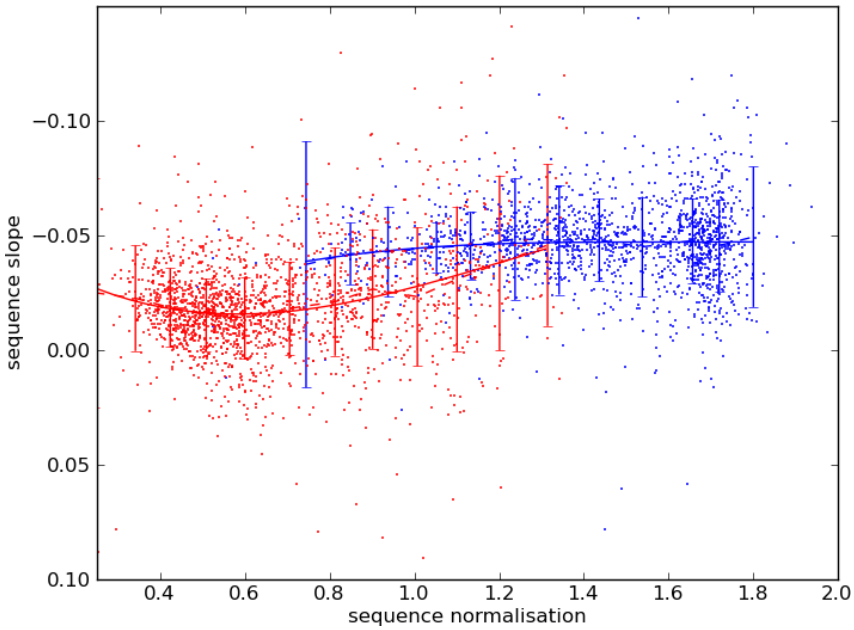

Figure 5.18: The sequence slope (β) evolution with colour normalisation $c_{\rm m20}$ in *g-r* (blue) and *r-i* (red) detected ORCA clusters from Stripe 82 data. The solid lines indicate splines fit to this cluster data, and are used to guide ORCA when searching for clusters in the DR7 dataset.

the splines and source data for *g-r* and *r-i* cluster detections. B-splines are completely defined by their knots, coefficients and polynomial degree. To facilitate the future recovery of these splines we provide this information in §C.6 of Appendix C.

5.6.5 The cluster catalogue

Our final step before applying ORCA to the survey involves setting data boundaries and areas for each cell with DARB . For cells at higher declinations, we found DARB underestimated the area compared to the HEALPix utility discussed later in §5.6.6. This arises from the difference in HEALPix projections and the conformal sampling grid applied to the data. We therefore opted to use HEALPix estimates for each cell area (with nside=128, see Table 5.5) but keep the DARB boundary and hole identifications. More details on the HEALPix utility can be found in the following section.

We apply the detection algorithm to each of the 152 cells, following which they are merged (and overlap regions processed) to form a final catalogue of 32,808 clusters with at least 5 members - 9,547 of these have $N_{\rm gal} \geq 10$. A total of 335,581 galaxies (1.4% of

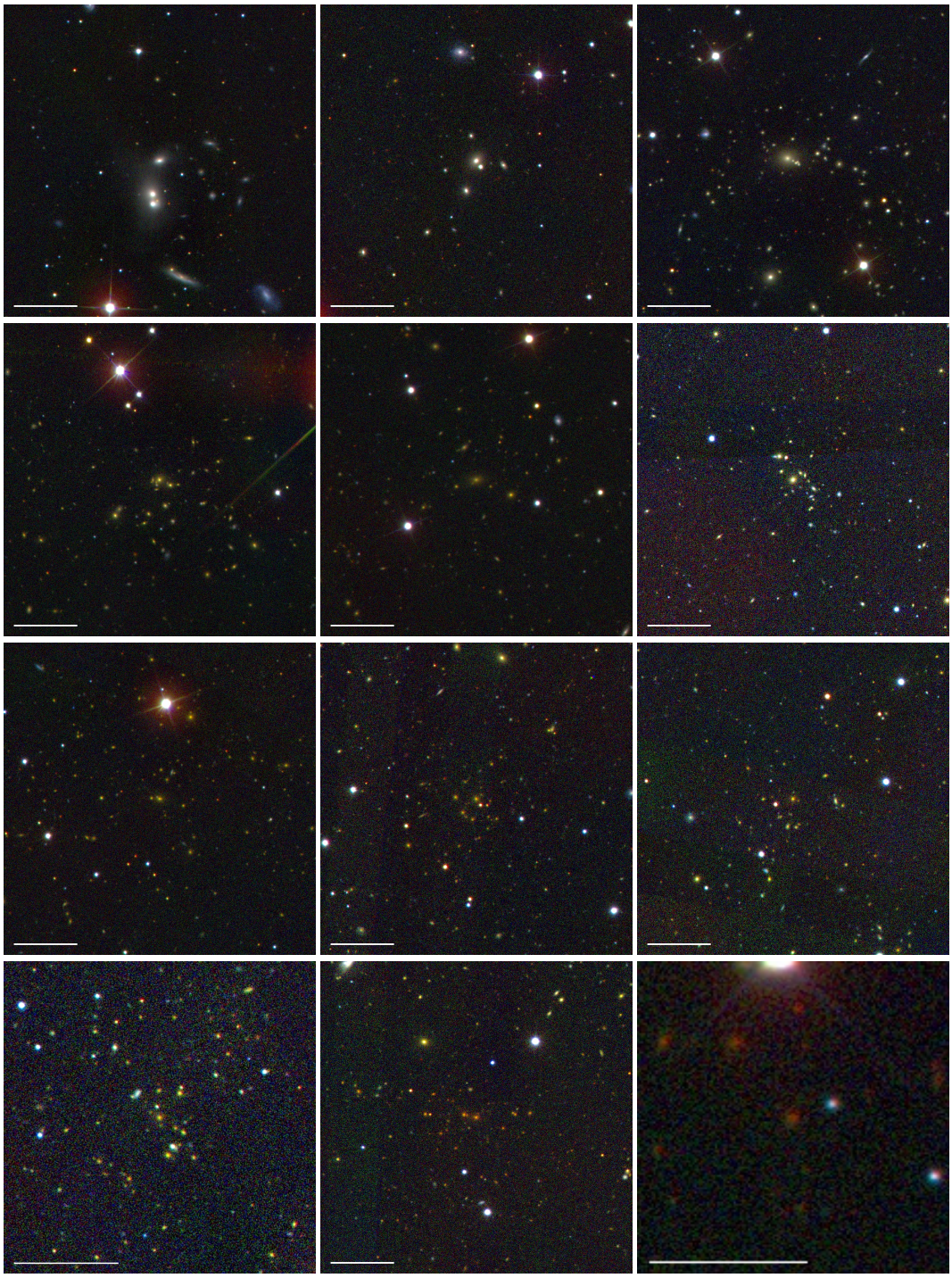

Figure 5.19: A gallery of cluster detections probing out to z=0.6. Top-left to bottom right shows a sequence of ORCA cluster detections from $z = 0.05$ to $z = 0.6$ in increments of $dz = 0.05$. All cluster redshift uncertainties are less than the sampling interval we use here, with the exception of the most distant cluster (with an uncertainty only marginally greater than the bin width). The horizontal white bar indicates a scale of 1′ in each image, which are predominantly 5′×5′.

the magnitude-limited source catalogue, excluding associate members) were assigned to ORCA clusters.

A gallery of example cluster detections from this catalogue can be seen in Figure 5.19, where in each case the image is centred on the cluster BCG. Clusters are ordered in ascending redshift in steps of $\mathrm{dz} = 0.05$ from top-left ($\mathrm{z} = 0.05$) to bottom right ($\mathrm{z} = 0.60$).

Cluster redshifts were calculated using the median-weighting prescription described in §3.4.2 by evaluating photometric redshifts (supplied for each galaxy) and available spectroscopic data. The median cluster redshift of the survey was $\mathrm{z_{med}} = 0.21$, but the cluster redshift distribution (the *top-left* panel in Figure 5.20), suggests cluster identifications out as far as $\mathrm{z} \sim 0.6$. However, when comparing the normalised redshift distribution to the Stripe 82 study (the *red* line in the *top-right* panel), we see this deeper data probes to considerably larger redshifts. In the *bottom* panel of Figure 5.20, we also plot a cluster density map indicating the regions of the survey where ORCA detected the greatest number of systems (*red*). The density of clusters drops with increasing declination (δ) because lines of constant RA converge when moving from the equatorial. This results in, by convention, a smaller sampling area in cells at high declinations (see the grey-scale density gradient in Figure 5.16) and reduces the number of clusters detected. Because in Figure 5.20 we plot clusters in a Lambert equal-area cylindrical projection, we weight each cluster by a factor $1/cos(\delta)$ when calculating counts in the histogram bins. For a statistical analysis of this cluster catalogue, detections should ideally be assigned weights according to their position in the survey, and could include higher weights for with high-declination systems.

5.6.6 Extrapolation to the 3π survey

In §5.4 we noted that Pan-STARRS SAS data is unrepresentative of the expected 3-year 3π data. In this section we extrapolate cluster yields from SDSS measurements under the assumption Pan-STARRS will attain at least SDSS single-epoch depth after three years of the 3π survey. To achieve this, a more precise estimate of the SDSS sky coverage is required, and for this we use the HEALPix[6] utility.

[6] http://healpix.jpl.nasa.gov

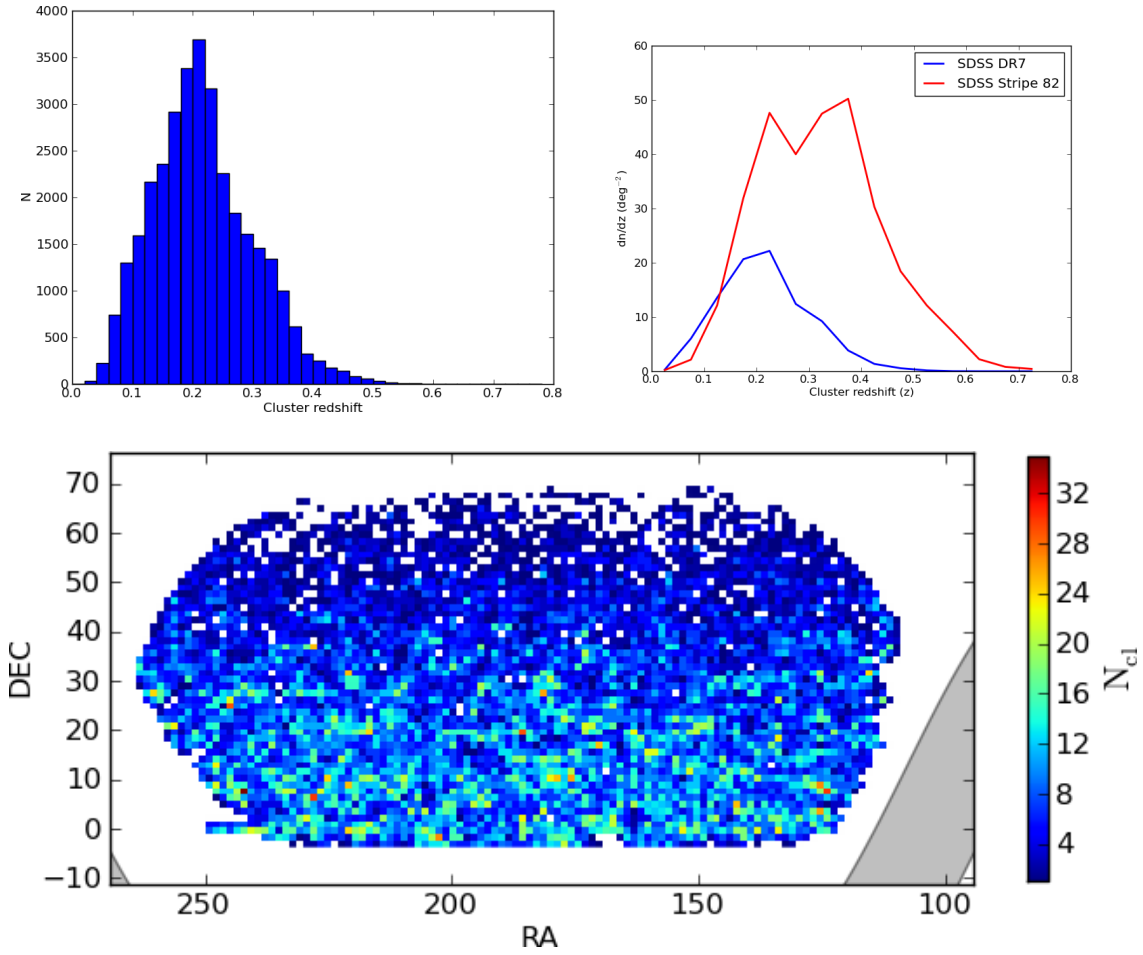

Figure 5.20: (*top*) The cluster redshift distribution of SDSS ORCA clusters, and a comparison to detections in the deeper Stripe 82 data. In the *left* plot, we show the ORCA DR7 N(z), whilst in the *right* plot we compare the area-normalised DR7 redshift distribution (blue) to that of the deeper Stripe 82 catalogue (red). (*bottom*) A two-dimensional histogram of the CR footprint showing the density of clusters ORCA detected. Redder colours indicate higher projected cluster densities, with each cluster assigned a weight $1/cos(\delta)$ to account for reduced completeness at higher declinations. The grey data at the edge of the plot are galactic latitudes $b \pm 10^{\circ}$ from the galactic plane.

nside	$N_{\rm cells}$	Resolution (′)
128	196608	27.5
256	786432	13.7
512	3145728	6.8
1024	12582912	3.4
2048	50331648	1.7

Table 5.5: HEALPix cell resolutions as the nside parameter varies. Care must be taken to select a cell size that neither oversamples nor undersamples the data.

HEALPix subdivides the sky into cells of equal area and uses a set of fast routines to calculate which cell a position resides in. These routines were used to count the total number of HEALPix cells occupied by a 10% random sample of the $\sim 122 \times 10^6$ galaxies in the CR footprint. Care must be taken to choose an appropriate HEALPix cell size, defined by the nside variable. Valid options for nside are shown in Table 5.5, along with the corresponding resolutions. We adopt nside=256 to generate cells not so small they oversample the typical inter-galaxy projected separation (leading to underestimates of the sky coverage), but not so large that significant coverage is claimed in the presence of only a few galaxies. A survey area of 7621deg^2 is calculated based on a HEALPix filling factor of 18.5% - this estimate is consistent with 7701deg^2 estimated from nside=128, and suggests a mean density of 4.3 clusters/deg^2 for the SDSS CR. Based on the redshifts ORCA is sensitive to, we probe a comoving volume exceeding $2h^{-3}\mathrm{Gpc}^3$.

To estimate the PS-1 extragalactic science footprint, we remove an area of $\pm 10^\circ$ in galactic latitude to account for the galactic plane - approximately 4200deg^2 assuming Pan-STARRS extends down to $\delta = -30^\circ$. We therefore expect galaxy detections over $\sim 26,739$deg^2. By achieving SDSS-depth imaging PS-1 will detect *at least* 1.15×10^5 clusters. This estimate does not account for y-band cluster detections in the 3π survey[7] or the reduction in extended source density at low galactic latitudes. Assuming one can detect clusters to the same redshifts as those identified in this study, ORCA clusters will cover a comoving volume of over $20h^{-3}\mathrm{Gpc}^3$. Of particular interest will be comparisons

[7]The depths explored here suggest this band will in any event by insensitive to cluster signals

of this predicted yield to projections from forthcoming mock PS-1 surveys.

5.6.7 Recovering large-scale structure with ORCA clusters

Adapting our filament-finding technique from chapter 2, we use the 3D positions of ORCA clusters detected in this survey to identify connected structures. After the initial identification of 2dFGRS groups and clusters, that particular study used a second Friends-Of-Friends algorithm on the positions of field galaxies and group centres. In this investigation, we used a simplified approach, applying the linking algorithm only to detected clusters. We estimate the ORCA cluster comoving space density in $25h^{-1}\mathrm{Mpc}$-width annuli and use this distribution to calculate the mean, redshift-dependent inter-cluster separation.

A spherical linking volume was used to connect ORCA clusters together, and the properties of the structures in this sample were investigated by varying the linking length b (preventing, as before, linking lengths exceeding $10h^{-1}\mathrm{Mpc}$). We adopt a value $b = 2.5$ times the mean inter-cluster separation, based on observed percolation characteristics analogous to those of the 2dFGRS study.

To calculate the luminosity of connected structures identified from cluster data, rest-frame SDSS and $\mathrm{b_J}$-band luminosities are determined with the `K-correct` code (Blanton and Roweis, 2007). Just as cluster luminosities are simply the summed luminosity of member galaxies, connected structure luminosities are the summed ORCA cluster luminosities. From ORCA cluster data, we detect 2421 connected structures (with at least two members) extending out to z=0.53 with a mean redshift $\bar{z}$=0.20.

Part of the SDSS data overlaps with the 2dFGRS NGP region; in Figure 5.21 we show a comparison of connected structures in this volume. In the ORCA (*left*) reconstruction, *grey* points indicate 2dFGRS galaxies from structures detected in our chapter 2 study with at least 10 members in addition to ORCA connected structures colour-coded by the number of members (*top*) and $\mathrm{b_J}$-band luminosity (*bottom*). Because in our DR7 study we only link ORCA clusters, the density of points is much lower than the 2dFGRS analysis, where each dot represents either a galaxy or a group centre. Nevertheless, it can be seen in both the membership (*top*) and luminosity (*bottom*) plots that ORCA clusters appear to reconstruct the *Sloan Great Wall* and trace the brighter systems at higher redshifts. The

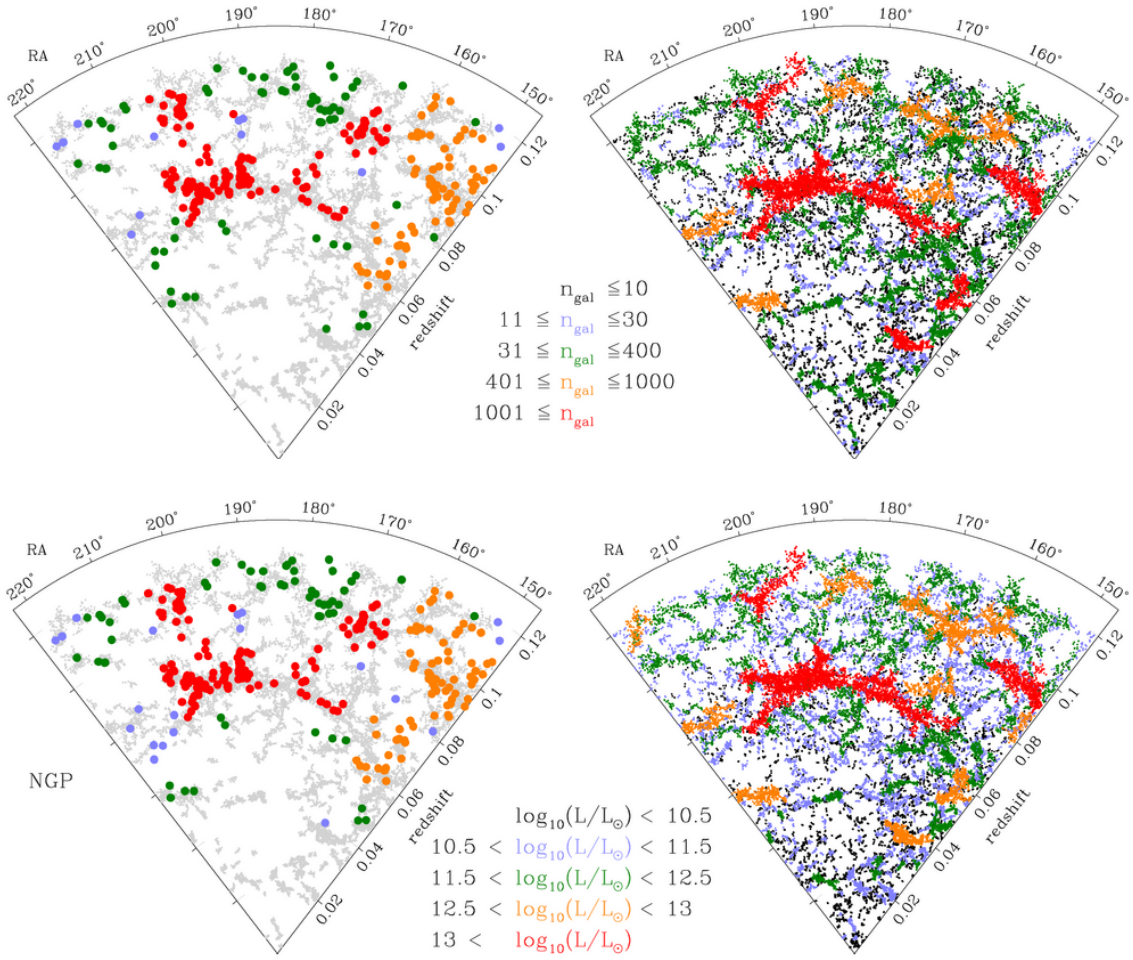

Figure 5.21: The reconstructed large-scale connected structure based on ORCA clusters (*left*) and spectroscopically measured galaxies in the 2dF-GRS (*right*; chapter 2). To aid comparison, grey data in the *left* panels are galaxies from spectroscopically-selected connected structures with at least 10 members. The *top* panels show the number of galaxies assigned to the connected systems, the *bottom* shows the K-corrected b_J-band luminosities.

ORCA connected system at z=0.11 and $\alpha \sim$205°, despite having a similar luminosity and membership to the *Sloan Great Wall*, is a separate system.

In chapter 3 (§3.7.1), we establish that ORCA detections suffer from low completeness at $z \leq 0.08$, explaining the few low-redshift structures identified in this survey. Moreover, SDSS data does not extend down as far in declination, suggesting some additional structures may also be missed. Nevertheless, we demonstrate that red-sequence techniques permit the reconstruction of large-scale structure previously identified with spectroscopy. This approach will enable future investigations at higher redshifts without spectroscopic measurements.

5.7 DXS SA22

In this section, we explore the detection of clusters in a band-merged optical/near-IR catalogue (J.W. Kim, priv. communication). In principle one may follow the 4000Å break out of the optical regime into the near-IR bands. This break straddles the *J* and *K* filters out as far as $z \sim 2$ and therefore allows (with sufficient depth) the exploration of clusters more distant than can be probed by optical data.

The SA22 field is centred at $(\alpha, \delta) = (334.40°, 0.32°)$ and has been the subject of previous analyses measuring the angular two-point correlation functions of Extremely Red Objects (ERO) and Distant Red Galaxies (DRG) (Kim et al., 2011). A catalogue of cluster galaxies would prove a valuable database to check, for example, the connection between environment and early-type galaxies. Moreover, with photometry spanning a wide spectral range, accurate photometric redshift estimates of the cluster members are possible.

The near-IR bands also provide good proxies for the cluster stellar mass. Near-IR luminosities are relatively insensitive to ongoing or recent star formation activity (Kodama and Bower, 2003), meaning *J* and *K*-band photometry could provide accurate estimates of galactic stellar mass. We illustrate this in Figure 5.22, where we show data from a mock survey (A. Merson, priv. communication) using the Bower et al. (2006) semi-analytic model over a $\sim$100deg^2 survey of galaxies out to $z \sim 4$. We calculate the observed-frame *r, J, K* luminosities and total (i.e. bulge + disk) stellar mass for all galaxies in haloes out to $z \leq 1.5$ with at least four other members. The total stellar mass and luminosity are then calculated for each halo - we plot these in Figure 5.22. As one

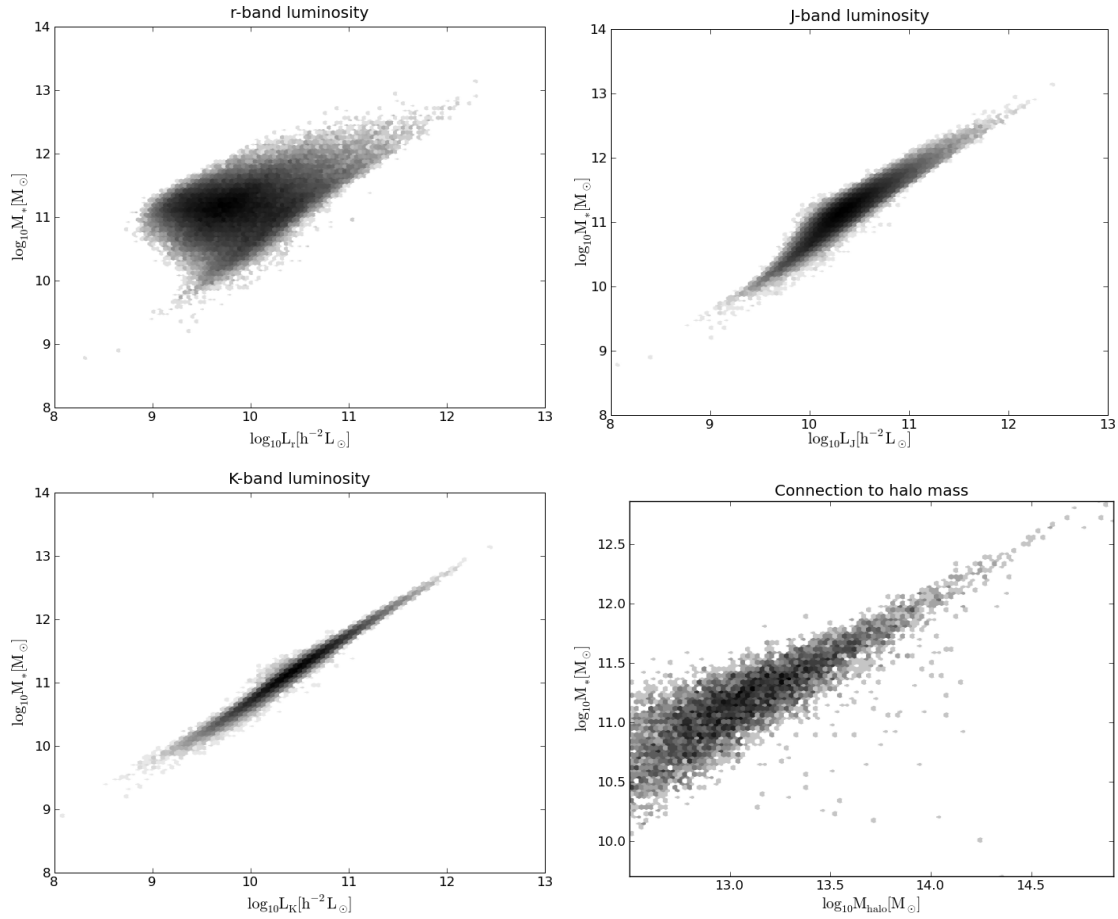

Figure 5.22: The correlation between observed-frame *r, J, K*-band luminosities and stellar mass in mock galaxy groups and clusters. As one observes with redder bands, the correlation becomes tighter. With an accurate handle on the cluster stellar mass, from near-IR photometry one may infer the mass of the underlying halo (*bottom-right*).

uses redder filters, the correlation between stellar mass and luminosity becomes tighter. Having a well constrained cluster stellar mass in turn provides a good estimate of the halo mass (*bottom-right* panel).

5.7.1 The data

The source data used in this study comprises *ugriz* Canada France Hawaii Telescope (CFHT) optical data and UKIRT Infrared Deep Sky Survey (UKIDSS) Deep Extragalactic Survey (DXS) *JK* 2″ aperture photometry covering $\sim$5deg^2 . Accompanying this catalogue are photometric redshifts generated with the EAZY (Brammer et al., 2008) pipeline, with 1-σ scatter in $\Delta z/1 + z_s$ of 0.05. To concentrate solely on the high-redshift cluster population in this survey, we will make use of the *i*, *J* and *K* bands only. Because the source data may feature drop-outs in the same manner described in §5.4 our analysis includes all galaxies with detectable magnitudes in at least two of the three bands.

We apply magnitude cuts to the survey to reduce the inclusion of galaxies with poor photometric accuracy, and follow the procedure laid out in §3.3.7 to derive limits of (i, J, K) = (25.2,21.2,23.6). These cuts resulted in an input catalogue of 303,008 galaxies with detectable magnitudes in at least two bands. This field has a number of excised regions corresponding to extended stellar haloes (the example field demonstrating DARB in Figure 5.1 is a subset of this survey). We apply DARB to these data, taking both excised regions and the survey boundary into account when calculating the survey area. DARB found the area covered was 5.3deg^2 and identified all excised regions in the data.

To search for high-redshift cluster sequences, we applied ORCA to this survey with the {*i-K*, *J-K*} and {*J-K*} filter combinations. Owing to the lower depth of the *J*-band, *i-K* was selected over *i-J*.

5.7.2 Photometric filter parameters

To establish the colour-magnitude slopes of the *i-K* and *J-K* sequences, we determine the predicted slope from the aforementioned 100deg^2 mock survey. We identify all haloes with redshifts $z \geq 0.8$, stellar masses over $10^{10.5} h^{-1} M_\odot$ and at least 5 galaxies. From this subset we estimate the median sequence slope in *i-K* vs *i* and *J-K* vs *K* between redshifts of 0.8 and 1.2. We found the slopes did not evolve over this range, and were β_{i-K}=-0.25

Colour	Slope (β)	Width (σ)	Range	1^{st}-stage filters
i-K	-0.25	0.152	$2.00 - 4.00$	51
J-K	-0.09	0.152	$0.66 - 1.70$	27

Table 5.6: Filter parameters used in the SA22/DXS `ORCA` detection scan.

and β_{J-K}=-0.09. This latter result agrees well with β_{J-K}=-0.088 measured for a z=0.89 cluster in Stott et al. (2009).

Stott (2007) note BCGs at z $\simeq$1 have colours of at least *J-K*=1.6 (Vega), and similarly use *I-K*=3.8 (Vega) as the basis for a red sequence search. Using the Blanton and Roweis (2007) `K-correct` conversions to AB magnitudes, we use these values to set the blue limits of our search. Red limits were established by visual inspection of the colour-magnitude diagrams.

Table 5.6 details these limits and summarises the parameters used in the `ORCA` sequence searches. For galaxies within 2.0<*i-K*<4.0 and 0.66<*J-K*<1.70, the average photometric redshift is $\bar{z}_{\rm p}$ =1.22$\pm$0.55, although we emphasise we do not use redshift data in the cluster detection process for this study.

5.7.3 Predicted cluster yield

Because the *J*-band is shallower than the others, it may possibly limit the detectability of clusters in the survey. We can predict both the yield and expected membership of cluster detections based on model data. An indication of our ability to select high-redshift cluster sequences is to determine how many magnitudes below M* the *J*-band reaches. The approach detailed here fits a Schechter function to the rest-frame *R*-band luminosity function in the z=0.98 Millennium Simulation snapshot (`snapnum`=41). This band corresponds to the wavelength region of the observed-frame *J*-band at z=0.98. From 500,000 galaxies selected from the Bower et al. (2006) catalogue in this snapshot, we derive *R*-band Schechter function parameters of $(\phi^\star, M_R^\star, \alpha)$=($3.8\times10^{-4}h^3{\rm Mpc}^{-3}, -20.85, -1.10$).

We select 100 galaxies with observed-frame *J*-band luminosities closest to $M_{\rm J}$=-20.64 (the flux limit *J*=21.2 at z=0.98) and use their rest-frame fluxes to find the characteristic rest *R*-band luminosity of observed-frame *J*-limited galaxies. From these galaxies, a limiting *R*-band magnitude of $R_{\rm lim}$=-18.33 was calculated: this is $\sim$2.5 magnitudes fainter

than $M_R^\star$. Returning to the Millennium Simulation, we may use this limit to establish the fraction of clusters detected in the z=0.98 snapshot. We retrieve galaxies with no Hα flux that belong to haloes with $\mathrm{M_{halo}} \geq 10^{14}h^{-1}M_\odot$. The simulation allocated this sample of 52,251 galaxies to 541 distinct dark matter haloes, with all haloes above the $N_{\rm gal}$=5-member limit for potential detection by ORCA (the smallest halo had 29 galaxies). We calculate that 235 of these haloes have $N_{\rm gal}$ $\geq$5 members brighter than our derived *R*-band detection limit, suggesting a completeness of 43%. Of those detectable haloes, the median number of galaxies brighter than the detection limit was 6, on average corresponding to 8% of the total galaxy clusters in the halo (the majority of these will of course be far below the detection limit).

As a rough cluster-yield calculation, if we take the projected area of the Millennium Simulation (with box length $500h^{-1}\mathrm{Mpc}$) to be $500\times500h^2\mathrm{Mpc}^{-2}$ at z=0.98, the 235 detectable clusters subtend 150.77deg^2 in the sky. This suggests ORCA should identify 8-9 clusters in the 5.3deg^2 SA22 footprint.

More precise measurements of the predicted cluster yield may come from light-cone mock simulation data such as the 7deg^2 Pan-STARRS MDS survey. This lightcone extends out to z=2.62 and was used in chapter 3 to determine the ORCA detection performance. From this mock data, we identify 24 clusters with $\mathrm{M_{halo}} \geq 10^{14}h^{-1}M_\odot$ and redshifts exceeding z=0.98. Assuming the detection completeness of 43% determined above, we calculate a cluster yield 9% higher than the Millennium Simulation estimate, corresponding to 9 clusters in the SA22 footprint. Because both mock datasets do not include photometric errors, they are likely to provide only upper limits to the detection yield from the SA22 field.

Before applying ORCA to the survey, we consider one final adjustment to the detection parameters. At larger redshifts, the reduced detectability of cluster members results in a lower cluster galaxy surface density compared to the $\mathrm{z} \lesssim 0.6$ detections we have been achieving with ORCA. Moreover, clusters at this redshift will not be as collapsed and indeed may yet be virialised. To quantify this we compare the 24 clusters identified in the mock lightcone survey above to all clusters $0.28 \leq \mathrm{z} \leq 0.32$ constructed from the same mock data in §3.6 of chapter 3. This lower-redshift sample comprised 64 clusters and brackets the median cluster redshift of the ΛCDM catalogue based on the SDSS Stripe 82 magnitude limits.

To estimate the cluster galaxy surface density for each cluster, we calculated the ratio of cluster members to the surface area of the cluster. This surface area was defined simply by πr_{max}^2, where r_{max} is the distance between the furthest galaxy and the cluster centre. To reflect our Millennium Simulation cluster recovery findings above, for the high-redshift sample, we randomly selected 8% of the members for each cluster and calculated the surface density from this subset. To ensure robustness against the random seed chosen, we repeated this random selection 200 times for each cluster. We find the low-redshift sample is on average 4.3$\times$ denser than the high-redshift sample. This will result in lower cluster-member Voronoi cell densities relative to the mean galaxy density for our high-redshift sample. Because the P_{thresh} parameter is measured relative to the mean density, one consequence of this will be fewer genuine cluster members considered part of the overdense cell population. We compensate for this effect by raising the probability threshold slightly to P_{thresh}=0.02.

5.7.4 The cluster catalogue

Following the application of ORCA to this galaxy catalogue, we detect a total of 54 galaxies assigned to 9 clusters, in line with our model projections. Of these, two detected in the {*i-K*, *J-K*} selection filter combination are considered high-significance detections, whilst the remaining seven were detected only in {*J-K*}. Figure 5.23 shows the position of each of the detected systems. Systems coloured blue are those selected in {*i-K*, *J-K*}, those coloured red are *J-K*-selected. The green lines indicate the survey boundary and excised regions detected by DARB. Using the photometric redshift data calculated for the cluster galaxies, we estimate a median cluster redshift of z_{median}=0.89. Table 5.7 shows the catalogue data for these detections. Because of the susceptibility of richness measures to Poisson counting errors, and the B_{gc} reliance on the redshift evolution of the luminosity function, it is likely the assumptions we made in §3.4.3 and §4.4.2 are no longer valid at the redshifts probed here.

The median number of galaxies assigned to clusters was 6, in agreement with the predictions of the model. Whilst the number of cluster detections matches well with our expected cluster yield, the models do not take into account any photometric uncertainties. It is therefore possible some of the low-significance clusters ORCA detected

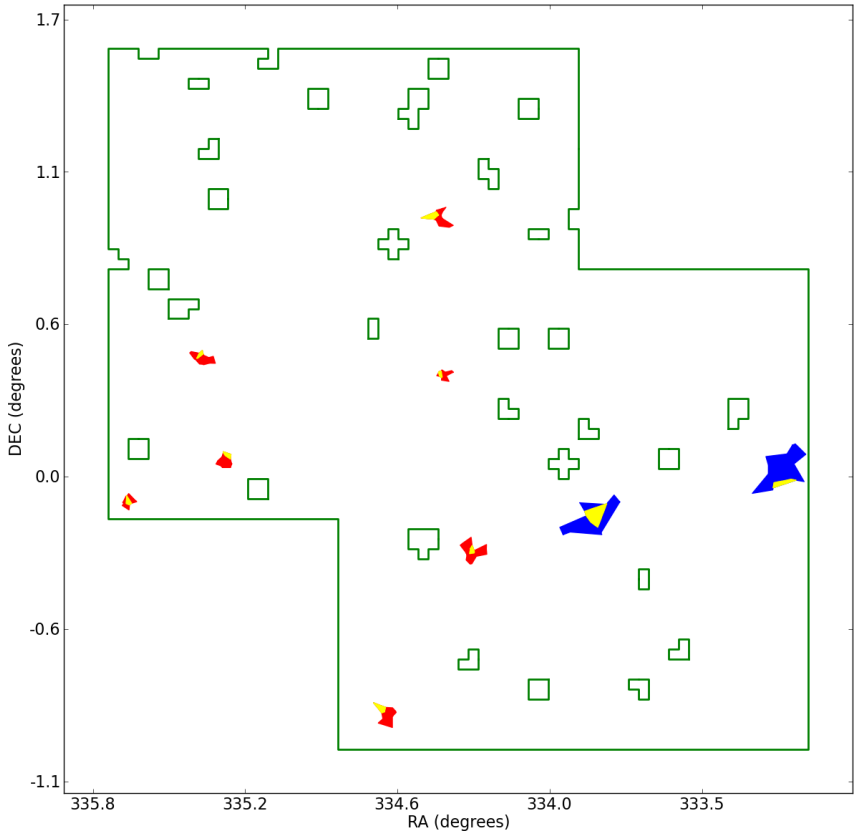

Figure 5.23: Clusters detected in the SA22/DXS. The coloured cells represent clusters identified in different colour pairs. Blue cells correspond to clusters detected in {*i*-*K*, *J*-*K*} filter pairs, red cells in {*J*-*K*} only. Yellow cells belong to the BCG of each cluster. The green lines delineate the survey boundary as estimated by DARB, in addition to the excised regions of the survey.

are spurious. We plan to assess these systems with UKIDSS and CFHT imaging. Of the two high-significance detections, the eastern cluster (5 members) has a photometric redshift of 0.97 ± 0.04, whilst the 6-member western cluster has a photometric redshift of 0.98 ± 0.04. We cannot exclude the possibility these may be part of the same structure. Both targets are therefore promising candidates for near-IR spectroscopic follow-up, and may reveal part of a larger, connected structure.

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	b_gc	scatter	θ_{80}	θ_{80}/θ_{20}
SA22 J221530-00090.1	333.87558	-0.15168	0.974	c0s0w0q0d0b0p5h0	5	-293	0.021	0.0012	1.929
SA22 J221239+00000.4	333.16282	0.00703	0.984	c0s0w0q0d0b0p6h0	6	-237	0.047	0.0011	3.429
SA22 J221718-00160.9	334.32541	-0.28170	0.869	c0s0w0q0d0b0p6h0	6	156	0.027	0.0004	1.636
SA22 J221837-00530.6	334.65357	-0.89388	0.943	c0s0w0q0d0b0p5h0	5	56	0.052	0.0003	3.000
SA22 J222121+00260.6	335.33810	0.44323	0.911	c0s0w0q0d0b0p10h0	10	-190	0.036	0.0004	2.111
SA22 J222229-00050.5	335.62123	-0.09096	0.818	c0s0w0q0d0b0p5h0	5	110	0.031	0.0003	13.500
SA22 J222101+00030.7	335.25384	0.06089	0.817	c0s0w0q0d0b0p6h0	6	183	0.045	0.0003	1.833
SA22 J221745+00220.0	334.43657	0.38259	0.863	c0s0w0q0d0b0p5h0	5	0	0.054	0.0003	2.625
SA22 J221749+00590.1	334.45273	0.98472	0.779	c0s0w0q0d0b0p6h0	6	0	0.043	0.0003	1.714

Table 5.7: Cluster data for detections in the SA22/DXS field

5.8 Summary

In this chapter we have demonstrated the application of ORCA to a variety of photometric surveys. Particularly important to the efficacy of our cluster detection is the accurate determination of survey area. To this end, we develop DARB - an extension to ORCA that calculates the area and boundary of the survey to be investigated. This extension also facilitates the process of cutting out data with the same footprint from another survey - this is utilised in our comparison of Pan-STARRS SAS data to SDSS data covering the same region. From our comparison we find that Pan-STARRS is not currently as deep as single-epoch SDSS coverage, but poor photometric conditions in test data of the former prevented a fair appraisal. In deeper MDS Pan-STARRS data, measurements from incomplete photometric redshift data and a spectroscopic sample of ORCA BCGs suggests more distant clusters are identified. From one of these MD fields, we calculate cluster stellar masses based on the y-band filter. These masses are found to correlate well with the number of galaxies ORCA identifies, but more work is required to improve the correlation of higher-level richness estimators with mass within ORCA . We describe the steps taken to extract cluster data from, to date, the largest available photometric dataset available - the SDSS DR7 survey. From this galaxy catalogue we detect 32,808 clusters that are subsequently used to identify the larger-scale connected structure in this volume. Part of this filamentary network coincides with the 2dFGRS study described in chapter 2, and structures common to both datasets are highlighted. Finally, we demonstrate how ORCA may be applied to near-IR data. Following an analysis of the expected yield and detectability of high-redshift clusters, we identify nine distant cluster candidates, two of which have redshifts close to z=1.

Chapter 6

Poisson point processes in Voronoi diagrams

6.1 Overview

In this chapter, we explore the fragmentation of two dimensional datasets into Voronoi cells, and how to characterise cells from a random distribution. A previous (widely used) study of this problem attempted to fit a model to randomly distributed data; we use the integral form of this model in chapter 3 when determining the statistical significance of cluster galaxy Voronoi cells relative to the field. As claimed by the original study, this model incorrectly predicts the cell area distribution in the high-density regime. It is of interest to investigate this in light of improved computational resources and the need to quantify the level of spurious signal from the chance clustering of random points. In §6.2 we introduce the original study, which we attempt to replicate in §6.3. In §6.4 we devise an improved experiment, and generate higher-resolution data. To improve our understanding of the high density random cell population, in §6.5 we fit alternative models to these new data.

6.2 The Kiang conjecture

Georgy Fedoseevich Voronoi studied the generalised n-dimensional case of what would later become known as the Voronoi diagram (Voronoi, 1908). It was in use prior to this however, famously applied when physician John Snow identified the source of London's 1854 Cholera outbreak by drawing the Voronoi cell of a water pump in Broad Street (now Broadwick Street). This geometrical technique has found modern use in a diverse range of fields (see Figure 6.1), from earthquake analysis (Schoenberg et al., 2009) to

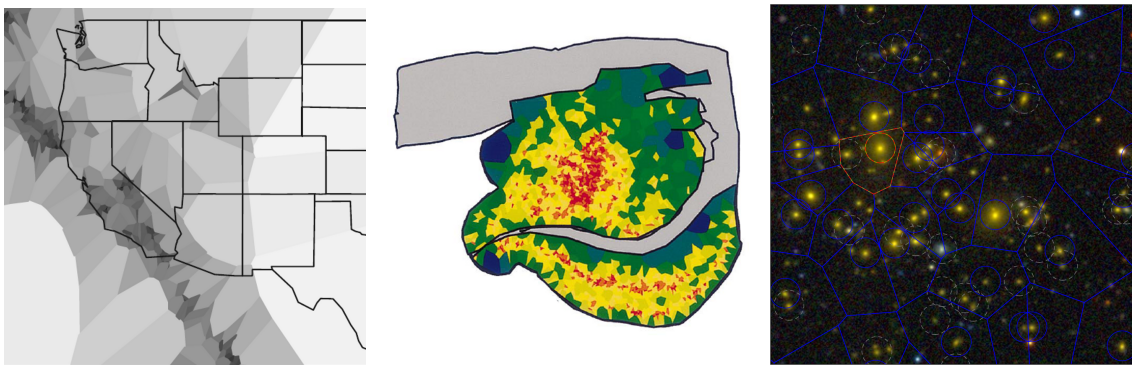

Figure 6.1: (*left*) A Voronoi estimator describing the spatial clustering of earthquakes on the west coast of the United States (adapted from Barr, 2008). (*middle*) Mapping the density of tagged neurones in a human amygdala - a medial temporal lobe of the brain (adapted from Duyckaerts and Godefroy, 2000). (*right*) The core of a galaxy cluster identified with the ORCA algorithm introduced in chapter 3.

mapping the distribution of (biological) cells in the Central Nervous System (Duyckaerts and Godefroy, 2000), and of course astronomy. The application of Voronoi diagrams to astronomical problems started with Kiang (1966).

A common thread running through many Voronoi-based studies exploits the connection between the clustering magnitude in a discrete point distribution and the area of their Voronoi cells. Moreover, because the Voronoi mesh and it's dual the Delaunay tessellation acts as an unbiased density estimator, it may be used to approximate a continuous density field (Schaap and van de Weygaert, 2000). In chapter 3, we discuss how best to evaluate the statistical significance of a projected concentration of galaxies sampled by a Voronoi diagram. Our approach, based on the Kiang (1966) study (hereafter K66) characterised the cell areas of a random distribution. The statistical properties of Poisson point processes can be used as a null hypothesis against which the measured distribution of a set of points can be compared. K66 used simple Poisson statistics to define the probability P(ds) that a 1D line broken up by randomly distributed points has a fragment length in the range [s,s+ds]:

$$\mathrm{P}(\mathrm{ds}) = \lambda \mathrm{e}^{-\lambda \mathrm{s}} \mathrm{ds}, \tag{6.1}$$

where λ is the mean number of random points per unit length. From this assumption,

Kiang derived the probability of recovering a Voronoi segment, the length between the midpoints of two random points (and the 1-D analogue of a Voronoi cell):

$$\mathrm{P(dx)} = 2\mathrm{x}e^{-\lambda 2\mathrm{x}}\mathrm{d(2x)}, \tag{6.2}$$

where x is the "standardised length": the segment length relative to the mean segment length. The partitioning of space in higher dimensions becomes harder to achieve objectively. The 2-D equivalent of break-points on a line are partition lines splitting the 2-D plane. Unlike the 1-D case, these partition lines intersect; the choice of their intersection points can be entirely arbitrary. To extend the segment analysis to higher dimensions, K66 hypothesised (later termed "Kiang's conjecture"; Moore and Angell, 1993) the generalised n-dimensional random partitioning of data into Voronoi segment lengths (1D), cell areas (2D) and polyhedral volumes (3D) followed a $\gamma(\mathrm{c})$-variate (with c=2n):

$$\mathrm{dP(x,c)} = \frac{\mathrm{c}}{\Gamma(\mathrm{c})}(\mathrm{cx})^{\mathrm{c}-1}e^{-\mathrm{cx}}\mathrm{dx}, \tag{6.3}$$

with x corresponding to the standardised lengths/areas/volumes relative to their mean values. K66 inferred a 2-D value of $c = 4.04 \pm 0.09$ from, at the time, a large Monte Carlo simulation comprising 112 realisations of a Poisson point-process. Despite claims refuting this conjecture (e.g. Sibson, 1980), rigorous 2 and 3-D derivation remains elusive and Equation 6.3 remains a good fit to the data.

6.3 Kiang comparison

We first compare the original K66 results to data from our attempts to replicate the original experiment. In what follows, we will adopt the K66 nomenclature. We refer to the randomly distributed data as "nuclei" - these are the data for which we seek cell areas. Nuclei are placed in a square box, populated by a regularly spaced lattice of "points". Lattice points are used in the K66 experiment to estimate the area belonging to each nucleus - more accurate area estimates are therefore possible with a higher density lattice (i.e. more points). In K66, 80 nuclei were randomly distributed in a square box populated by 6400 lattice points (a grid of 80×80).

Lattice points were then assigned to the closest nucleus; the area of each nucleus (relative to the mean lattice density: 6400 lattice points / 80 nuclei = 80) was approximated

by the number of allocated points. This process, repeated with 112 different random distributions, generated the distribution of normalised areas the model in Eqn 6.3 was fit to. In Figure 6.2, we show the distribution of standardised areas from the original study (*black*) and the dP(x,c=4) model fit (*red*) proposed by K66. This Figure reinforces remarks in K66 that highlight the failure of the model to adequately fit the data at $x < 0.8$, particularly $x < 0.25$ (*inset*). The high density domain is most relevant to the identification of clustered signals, where cell areas should be smaller than the mean area. The vertical dotted line in the Figure, corresponding to P(x,c=4)=0.01, indicates the largest area ORCA selects as a potential cluster member in chapters 3-5. Because the K66 model underpredicts the number of high-density cells arising from a random distribution, only a lower-bound on the predicted level of contamination from the field galaxy population is possible[1].

To reproduce these results, we repeat the above procedure using the exact same number of randomly distributed nuclei inside boxes containing the same number of lattice points. Because nuclei near the box edge will not be assigned as many lattice points as more central nuclei, we produce periodic replications of both the random nuclei and the lattice grid to surround the central box. We measure only nuclei in the central box, but those close to the edge are able to gain lattice points from beyond the box boundary. Conversely, nuclei from outside may also be assigned points from within the central box. We illustrate this arrangement in Figure 6.3 - blue cells measured in the experiment are surrounded by red cells from the periodic replications. Although this implies the sum of lattice points assigned to measured nuclei is not necessarily 6400, tests from our 112 runs revealed the mean number of points assigned per box is 6400 ± 2. We show the cell area distribution from our study as the blue dashed line in Figure 6.2; uncertainties in these data are calculated from Poisson counting errors in each bin. This distribution is a good match to the original K66 experiment, and again highlights the model's underestimation of high-density cells.

[1] We make the assumption here that field galaxies are randomly distributed; in reality they are less clustered than early-type galaxies.

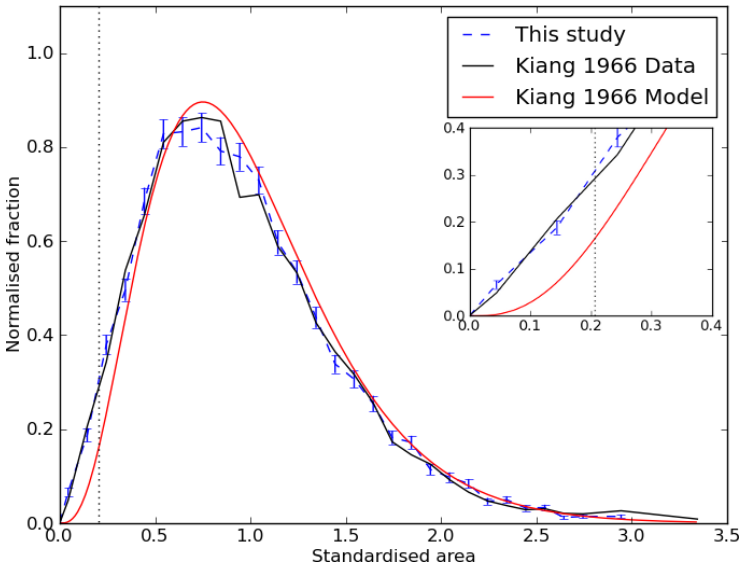

Figure 6.2: Nuclei area distributions from the K66 study and our attempt to recreate these data. The black line indicates the nucleus area distribution as recorded in K66, whilst the red line shows their model fit. The blue dashed line indicates our repetition of the original experiment, with uncertainties based on Poisson errors. The vertical dotted line notes the standardised area corresponding to P_{thresh}=0.01 in the ORCA algorithm.

6.4 Improved approach and resolution

There are broadly two approaches to measuring the cell area distribution. In the previous section, we indirectly infer the Voronoi cell areas of nuclei by determining the number of lattice points closest to them. This "statistical" approach does not however explicitly calculate cell areas, and may introduce a source of error arising from inaccurate area estimates. In this section, we use an "exact" method to define the boundaries of Voronoi cells belonging to the nuclei.

The `qhull` utility (Barber et al., 1996) was used to calculate the convex hulls (Voronoi cells) of all nuclei and determine the area of each cell. This method allows more precise estimates for smaller cell areas, where scales below the lattice sampling frequency potentially lead to area underestimates. Figure 6.3 (referred to above) shows the `qhull`-calculated cell boundaries for an example dataset. To account for cells extending over the central box boundary, we define the mean area in each run as the ratio of summed

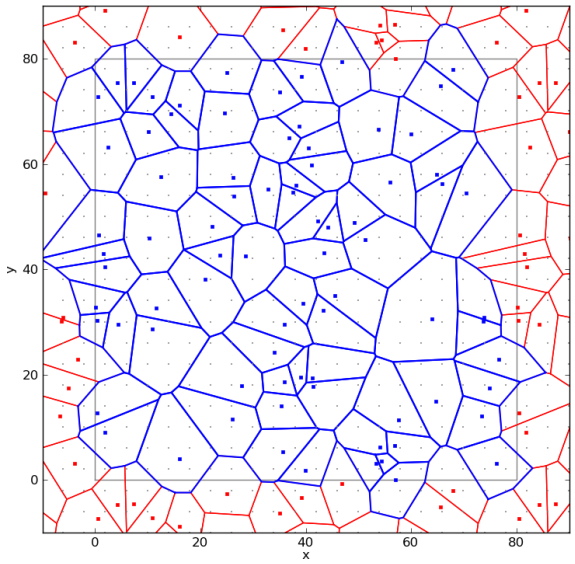

Figure 6.3: Example of a Voronoi diagram analysed in this study. The central box is delineated by the grey square. Blue and red points are the randomly distributed nuclei, the latter being periodic replications of data in the central box. The small grey dots (one quarter the density of those actually used in the analysis) are the lattice points used as estimates for the area belonging to the nuclei. The cell boundaries depicted here, generated by `qhull`, play no part in the analysis at this stage, but were included to improve clarity.

cell area to the number of nuclei in the box.

To compare the exact and statistical methods directly, in Figure 6.4 we plot the cell area distributions for the two methods (using the same number of nuclei), demonstrating the two approaches are broadly similar. Closer inspection of the high-density regime (*inset*) reveals `qhull` (*solid blue*) systematically predicts more high-density cells, in line with our expectations of higher sensitivity from this approach.

Having verified our "exact" method produces a similar distribution to the lattice-based method, to increase the range of standardised areas probed we next generate a new set of data with a greater number of nuclei. We produce 3750 Monte Carlo runs with 80 nuclei per box (corresponding to 3×10^5 Voronoi cells in total). After establishing the cell area distributions of these data are in line with the measurements taken thus far, we next fit alternative models to this new, high-resolution dataset.

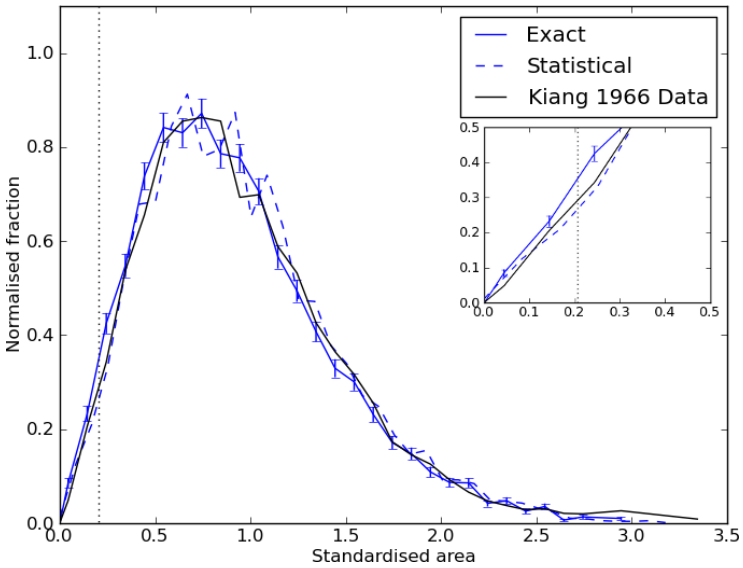

Figure 6.4: A comparison of two techniques used to generate the Voronoi cell area distribution for Poisson data. The blue dashed line indicates the data produced in §6.3, whilst the blue solid line is the more precise `qhull` approach detailed in §6.4. Although the distributions are similar, the high-density regime (*inset*) shows the "exact" approach predicts more small-area cells. As before, the black line indicates the original K66 data, and the vertical dotted line the standardised area corresponding to the `ORCA` selection threshold.

6.5 Alternatives to the K66 model

Whilst in this section we do not provide a rigorously-derived model of the 2-D cell area distribution, we do hope to improve on the K66 description of Voronoi cell areas, especially in the high density regime. We propose fits to the data with two types of model. Model one is an exponential-type function:

$$dP(\mathrm{x},\mathrm{a},\mathrm{b},\mathrm{c}) = \frac{\mathrm{xae}^{-\mathrm{bx}^{\mathrm{c}}}}{\sqrt{\mathrm{bc}}}\mathrm{dx}. \tag{6.4}$$

Model two is a more generalised form of the original K66 distribution described in Eqn 6.2:

$$dP(\mathrm{x},\mathrm{a},\mathrm{b},\mathrm{c}) = \frac{\mathrm{x}^{\mathrm{a}-1}\mathrm{b}^{\mathrm{a/c}}\mathrm{e}^{-\mathrm{bx}^{\mathrm{c}}}}{\Gamma(\mathrm{a/c})}\mathrm{dx}, \tag{6.5}$$

where a, b & c are free parameters used to derive the best χ^2-fit to the data; we note both models are not normalised.

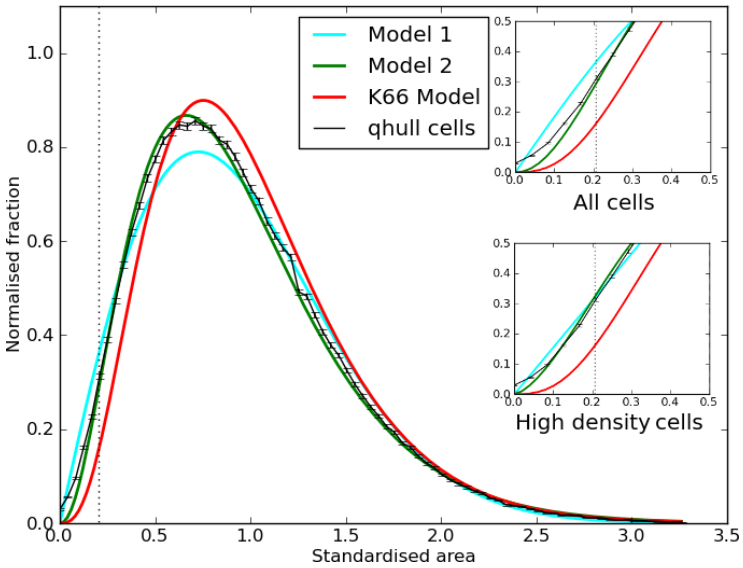

Figure 6.5: Model fits to the full set of Voronoi cell areas generated by `qhull` (black line). Error bars in these data arise from Poisson counting errors in each bin. The blue, green and red lines indicate model fits referred to in the text. The two inset panels show these fits in the high density ($x \leq 0.5$) part of the distribution (*top*) and for fits specifically to this domain (*bottom*). The best-fit parameters for these models can be seen in Table 6.1. As before, vertical dotted line notes the standardised area corresponding to $\mathrm{P}_{\mathrm{thresh}}$=0.01 in the `ORCA` algorithm.

In our analysis, the models are fit to the data in three different ways. The first approach considers all data generated, the second fits only to the "high density" ($x \leq 0.5$) part of the distribution, and the third focuses on the "low density" ($1.5 \leq x \leq 2.5$) domain. A fit to the low-density region of the data may be useful for the identification of cosmological voids.

In Figure 6.5, we show the best model fits to the whole distribution. The *black* line indicates the cell area distributions from the 3×10^5 Voronoi cells. We increase the number of area bins from 30 to 80 to allow greater sampling sensitivity; error-bars indicate the Poisson error in each bin from the 3750 runs. The *red*, *blue* and *green* lines indicate fits from the K66 (Eqn. 6.3), exponential (Eqn. 6.4) and three-parameter gamma (Eqn. 6.5) models respectively. Although both new models appear to provide a better account of the high-density region than the K66 study (see *top-inset*) the quality of fit justifies

Model	Full distribution	χ^2
1	a=2.50, b=0.97, c=1.86	3145
2	a=3.26, b=3.38, c=1.01	1657
K66	a=4.04, b=1.0, c=1.00	8423
Model	High density ($x \leq 0.5$)	χ^2
1	a=4.42$\times 10^{-3}$, b=0.84, c=1.81$\times 10^{-6}$	751
2	a=2.53, b=2.68, c=1.91	603
K66	a=4.04, b=1.0, c=1.00	6863
Model	Low density ($1.5 \leq x \leq 2.5$)	χ^2
1	a=4.55, b=1.46, c=1.48	13
2	a=3.50, b=2.89, c=1.16	13
K66	a=4.04, b=1.0, c=1.00	291

Table 6.1: Best-fit parameters to the high-resolution Voronoi cell data. For each model we quote the best-fit parameters and χ^2 to the whole dataset (columns two and three).

our separate analysis of particular portions of the data (as seen the *bottom-inset* panel showing the fit to only high density cells). In Table 6.1 we show the best-fit parameters for the models shown in this Figure (*top*), and for fits to the high density (*middle*) and low density (*bottom*) portions of the distribution. Model 2, the modified K66 distribution, appears to best fit the data in all three tests, and represents a significant improvement over the original model.

In our DXS/SA22 analysis of the previous chapter (§5.7), we attempted to compensate for the lower surface density of high-redshift cluster galaxies by raising the $\mathrm{P_{thresh}}$ parameter, thus including cells with larger areas. If we had an analytic solution to the distribution of random 2-D Voronoi cells, the $\mathrm{P_{thresh}}$ parameter would correspond exactly to the fraction of cells in this sample with areas smaller than x for P(x) = $\mathrm{P_{thresh}}$. Not only does no such solution exist, but as we have demonstrated, our fits to the cell area distribution are inexact. `ORCA` uses the K66 model to analyse the significance of each galaxy cell. Because we know this model underestimates the number of high-density cells, the K66 $\mathrm{P_{thresh}}$ will not select the same galaxies that an identical threshold would

for Model 2[2]. Using the improved fit to random data developed in this chapter, it is of interest to compare predicted fractions of cells below the threshold for the two models. For each $\mathrm{P_{thresh}}$ in Table 6.2, we determine the standardised area cumulative frequency based on Eqn 3.1 for K66 and:

$$\int dP(\mathrm{x,a,b,c})\mathrm{dx} = \int \frac{\mathrm{x^{a-1}b^{a/c}e^{-bx^c}}}{\Gamma(\mathrm{a/c})}\,\mathrm{dx} = -\frac{\Gamma_\mathrm{i}(\mathrm{a/c,bx^c})}{\mathrm{c}\Gamma(\mathrm{a/c})} + \mathrm{C}, \tag{6.6}$$

for the high density fit to Model 2, where Γ and Γ_i respectively denote the gamma and incomplete gamma functions. The constant of integration, C, is determined from the boundary condition P(0,a,b,c) = 0. Although we note Model 2 is not normalised, it remains a valid fit to the Voronoi cell data for the domain $x \leq 0.5$. In column two of Table 6.2 we determine the standardised area corresponding to each $\mathrm{P_{thresh}}$ from the $3{\times}10^5$ Voronoi cells in our sample; these areas are used to evaluate the cumulative fractions from both models. Columns three and four show $f_{model} = P_{model}(x)/\mathrm{P_{thresh}}$: the predicted fraction of cells below the threshold compared to the actual number.

For ORCA , our choice of $\mathrm{P_{thresh}}$ was observationally motivated - in §9 we adopted the threshold best recovering a population of cluster galaxies that had been visually identified. The threshold choice was further influenced by comparing the size of the recovered system to the characteristic $\sim 1h^{-1}\mathrm{Mpc}$ galaxy cluster scale at that redshift. In this sense, $\mathrm{P_{thresh}}$ merely acts as a tuning dial: given the known inaccuracies in the model, values do not describe the statistical significance of the cells. Using the data in Table 6.2, one can calibrate this dial to determine the actual fraction of spurious cells.

With $\mathrm{P_{thresh}} = 0.01$, K66 identifies $(1/0.2434){\sim}4\times$ too many overdense cells relative to the data, placing the true K66 threshold closer to $\mathrm{P_{thresh}}$=0.04. To recover the same population of cells using Model 2, one must adopt $\mathrm{P_{thresh}}$=0.05 (0.05/1.2627∼0.04). Similarly, the $\mathrm{P_{thresh}}$=0.02 equates to a true threshold of $\mathrm{P_{thresh}}$=0.053. It is important to note, within this context of selecting a threshold, there is not strictly a "better" model - K66 and Model 2 merely provide mappings to the standardised area. Whilst the analysis performed here permits a better insight into the spurious cell fraction, by design it does not empower us to improve on it, irrespective of model.

An entirely alternative approach to the choice of threshold would be to decide *a priori* the acceptable level of contamination within the cluster catalogue. If one were to adopt

[2]Equivalently, the threshold in standardised areas will be different between the two models.

P_{thresh}	x	f_{K66}	f_{Model2}	P_{cl}
0.005	0.0990	0.1496	1.0239	4.0×10^{-09}
0.010	0.1370	0.2434	1.1479	1.3×10^{-07}
0.015	0.1660	0.3194	1.2262	9.8×10^{-07}
0.020	0.1890	0.3747	1.2602	4.1×10^{-06}
0.025	0.2080	0.4145	1.2690	1.3×10^{-05}
0.030	0.2250	0.4486	1.2749	3.1×10^{-05}
0.035	0.2400	0.4752	1.2724	6.8×10^{-05}
0.040	0.2540	0.4996	1.2711	1.3×10^{-04}
0.045	0.2670	0.5210	1.2683	2.4×10^{-04}
0.050	0.2790	0.5388	1.2627	4.1×10^{-04}
0.055	0.2900	0.5528	1.2536	6.5×10^{-04}
0.060	0.3010	0.5686	1.2499	1.0×10^{-03}

Table 6.2: The accuracy of model predictions for the fraction of overdense cells below the range of thresholds P_{thresh}. The final column shows the probability a cell is a member of a spurious cluster with 4 other overdense neighbours.

a 1% limit, then it is clear Model 2 provides a much better estimate of the true contamination level. This is particularly useful for deciding the threshold to adopt in order to include not more than N spurious cluster detections in a large survey. The drawback to this approach is primarily motivating the (potentially arbitrary) choice of threshold. Nevertheless, this alternative may be of great benefit in studies where the level of permissible spurious noise is specified by the required sensitivity of the experiment.

It is therefore of particular interest to determine the number of spurious clusters produced in a survey. Within the ORCA prescription a spurious cluster consists of $N_{min} \geq 5$ field galaxy cells below P_{thresh}, each sharing at least one vertex with another member. Rigorous estimation of this rate, not attempted here, involves the creation of new Voronoi diagrams from two-component, clustered point-processes. Efforts to model this require a detailed understanding of the clustering statistics for the two components (field and

cluster), and a means to produce the distribution[3]. The simplified case we present here - namely that of a random field, estimates the spurious cluster rate arising from a purely Poisson process. If we make the simplifying (and perhaps naive) assumption that cell areas are independent of their neighbours a formulation for the probability a random cell is part of a spurious cluster with the lowest possible membership would be:

$$\mathrm{P_{cl}} = \mathrm{P_{thresh}}(\mathrm{N_c P_{thresh}})^{\mathrm{N_{min}}-1}, \tag{6.7}$$

where N_c is the number of neighbouring cells and N_{min} is the minimum number of connected cells ORCA classes as a cluster. Both analytical and computational studies (see e.g Tanemura, 2003) suggest $\langle N_c \rangle = 6$, and we use the ORCA value $N_{min} = 5$. P_{cl} estimates are shown in column five of Table 6.2 for each of the thresholds investigated. As anticipated, the likelihood of producing a cluster from random cells rises with threshold. In adopting a (K66) $\mathrm{P_{thresh}}$=0.02 in the aforementioned DXS/SA22 analysis, the chance of producing a spurious cluster increases by a factor of 5. From the 63 Voronoi diagrams analysed during searches through DXS/SA22 colour-magnitude spaces, 10% had sufficiently large cell counts to produce one spurious cluster. Under our ORCA prescription, we mitigate these false detections of random-cell clusters in two ways. Firstly, photometric selection filters in two colours reduce the number of cells in the field. Secondly, ORCA prevents the selection of systems below a critical density described by the Σ_{crit} parameter in §9. Both of these measures should be sufficient to keep the false detection rate low.

6.6 Summary

In this chapter, we have investigated the distribution of standardised (i.e. relative to the mean) Voronoi cell areas for a random distribution. We recreate the experiment originally used to derive a model fit to this distribution, and verify the model under-predicts the number of high-density random cells. This is of particular relevance to the detection of cluster galaxies with Voronoi diagrams, as these cells will also have high densities. We improve on the original experiment by generating Voronoi diagrams comprising 3×10^5 cells, calculating each Voronoi cell explicitly. After fitting two models to these data, we

[3]The Soneira and Peebles (1978) hierarchical clustering recipe is one credible method.

find a three-parameter gamma model best describes both the overall distribution and separate fits to the high and low density portions. We use this model to provide a better estimate of noise arising from random cells when choosing the $\mathrm{P}_{\mathrm{thresh}}$ parameter for ORCA.

Chapter 7 | *Summary*

7.1 Overview

In this thesis, we develop methods to identify galaxy groups, clusters and filaments in an evolving era of astronomy. Where accurate 3D spatial galaxy data are available, we provide a prescription to extract the connected structure of the cosmic web. Because the properties of systems detected depend on the survey geometry, the most appropriate interpretation of the data is through comparison with mock catalogues. Despite broad similarities between the two, it remains a challenge for models to reproduce the most extreme structures observed. We demonstrate that photometric redshift estimators currently lack the required accuracy to faithfully reconstruct the cosmic web from wide-area photometric surveys. As an alternative approach, we next use multi-band photometric data to exploit the bimodality of galaxy colours, constructing an algorithm that searches for clusters of galaxies. We quantify the detection performance of this algorithm with a mock survey and subsequently apply it to a range of real galaxy surveys. In the largest cluster census, we detect some 34,000 clusters over 7,600deg^2. The three-dimensional positions of these clusters are used to reconstruct the cosmic web within a $3h^{-3}\mathrm{Gpc}^3$ comoving volume identifying, where the data overlap, structures also detected in the spectroscopic 2dFGRS (including the *Sloan Great Wall*). Development of the detection algorithm is ongoing, and we highlight one promising method to better characterise the projected galaxy overdensity identified by both this implementation and similar approaches.

7.2 Key findings

In chapter 2 we introduce a prescription to identify connected systems of galaxies detected in spectroscopic surveys. The key result of this work is the broad agreement between properties of observed and model structures, except at the lowest and highest luminosities probed. These discrepancies can be traced to inadequacies in the galaxy formation model used. In both mock and real surveys we find the majority of connected systems arise from two-member L* galaxies. However, only in the 2dFGRS are filamentary systems of up to nearly $L_{\mathrm{b_J}} \sim 10^{14}\,h^{-2}\,\mathrm{L}_\odot$ observed, including part of a large structure known as the *Sloan Great Wall*. To compensate for imperfections in the model we apply a series of luminosity cuts to both mock and real input galaxies, but still find mock surveys do not generate enough high-luminosity systems. We next considered a scenario where the 2dFGRS survey had access solely to photometric data. Taking cues from our analysis of photometric redshifts detailed in chapter 3, we simulated the recovery of 2dFGRS structures under varying levels of redshift uncertainty. In order to recover the number of median-luminosity 2dF structures identified with spectroscopic data, photoz estimators must reduce their redshift uncertainties by *at least* an order of magnitude.

In chapter 3 we take an alternative approach to the analysis of multi-band surveys. Blind, systematic searches for "red sequences" in colour-magnitude space are coupled to Voronoi-based projected density field estimates with an algorithm named ORCA: an automated prescription for identifying galaxy clusters. Using a mock Pan-STARRS MDS galaxy catalogue, we compare ORCA-detected clusters to those inferred from the assignment of galaxies to dark matter haloes. ORCA reliably detects clusters in halo masses down to $10^{13}h^{-1}M_\odot$ and out to $\mathrm{z} \sim 0.6$. Relative to the model clusters, ORCA detections are 75% complete down to $M_{\mathrm{halo}} = 10^{13.4}h^{-1}M_\odot$ and we recover at least 75% of the cluster stellar mass content down to $M_{\mathrm{halo}} = 10^{13.8}h^{-1}M_\odot$. This latter point suggests, with supporting near-IR photometry, ORCA could be used to reliably estimate the mass of dark matter haloes hosting galaxy clusters. Tests with real astronomical data are consistent with findings from mock surveys showing the spurious detection rate is below 1%.

In chapter 4 we describe developments to the algorithm required to process large

volumes of information in wide-area astronomical surveys. Improvements to the primary computational bottleneck, the percolation routine, resulted in an order of growth evolving as $\mathcal{O}(\mathrm{N}^{1.77})$, and an algorithm approximately 3 times faster than earlier efforts. This reduced the level of spatial partitioning required when processing large-area surveys. We present the first "large-survey" application of the ORCA detector by searching for clusters in the deep 270deg^2 SDSS equatorial Stripe 82. From this data we detect over 4,000 clusters. Based on available photometric and spectroscopic redshift estimates, we derive a median cluster redshift of $\mathrm{z_{med}} = 0.32$. Significantly, we also determine a BCG photometric redshift uncertainty $\sigma_{\delta_\mathrm{z}} = 0.017$ compared to available spectroscopic redshifts. This level of accuracy suggests cluster redshifts can be confidently determined if BCG photometric redshift data are available. Without such data, one can still empirically fit clusters to a colour-colour track derived from detections made in this study, providing an approximate measure of their redshift. We find the BCG colour evolution is well matched to predictions of model LRG colours. Cluster members are monotonically bluer relative to this model as we probe fainter down the cluster red sequence, and we attribute this effect to a mass-metallicity rather than age-driven process.

Continuing our programme to identify galaxy clusters using the red-sequence technique, in chapter 5 we describe current progress in a series of photometric surveys, including samples from the Pan-STARRS project. We compare data representative of the three-year 3π-str PS-1 survey to single-epoch SDSS data, finding the latter probes deeper in redshift and recovers more clusters. This arises from unexpectedly poor photometric conditions the PS-1 data was taken in, and future photometry should be at least as deep as the SDSS. Spectroscopic confirmation of selected cluster BCGs in example (but not final-depth) PS-1 Medium Deep Survey data confirms we are able to detect clusters out to $\mathrm{z} = 0.5$ in deeper exposures with the Pan-STARRS Gigapixel camera. Using y-band data from a Medium Deep field, we are able to make estimates of cluster stellar masses. These masses are found to correlate well with the number of galaxies that ORCA detects. Extending the techniques developed to analyse large surveys in chapter 5, a 7,600deg^2 census of SDSS data generates a catalogue of over 32,000 groups and clusters with at least 5 members. Although a few clusters can be found at redshifts also probed by deeper Stripe 82 data, the median cluster redshift is considerably lower (0.21 vs. 0.32). We take this 3D dataset and link ORCA clusters together to identify connected

structures. Comparing our findings in a region common to both this data and the 2dFGRS, we demonstrate this approach can be used to identify connected systems down to $L = 10^{13}\,h^{-2}\,\mathrm{L}_{\odot}$, including the *Sloan Great Wall*. Extrapolating our findings to the area covered by the Pan-STARRS 3π survey, we expect to recover over 1.15×10^5 clusters based on an SDSS-depth in the *griz* bands. We then search for clusters at redshifts where the 4000Å break moves into the near-IR. With a combination of deep optical and near-IR data covering 5.3deg^2 we identify two high-probability detections at $\mathrm{z_{photo}}$ $\sim$1 and seven lower-significance detections selected in *J-K* only. This detection yield matches that predicted from model data.

Finally, in chapter 6 we discuss improvements to the analysis of randomly distributed datasets with Voronoi Diagrams. We investigate the work of Kiang (1966) within the context of cluster detection, and suggest an improved fit to the normalised cell area distribution better characterising the high-density region. We discuss the implications of this new fit, and provide estimates of the spurious cluster rate. Adoption of the new model should improve the separation of clustered data from random data.

7.3 Plans for future research

Much of the work detailed in this thesis has been developmental: producing and testing tools to generate filament or cluster datasets. Working from these efforts, we hope to enter the exploitation-stage of research via three themes:

- Producing science from current or planned galaxy cluster catalogues
- Refining and extending detection techniques to improve the quality of galaxy cluster science
- Developing the tools and facilities to process, distribute and explore vast cluster galaxy datasets

7.3.1 Science exploitation of ORCA clusters

Unveiling the y-band red sequence

Sandwiched between well characterised cluster catalogues out to $\mathrm{z} = 1$ and the emerg-

ing population of vigorously star-forming protoclusters at $z > 2$ is a region in which cosmological dimming and the lack of star formation conspire to form an observational impasse and a dearth of confirmed clusters. In this "cluster desert" a critical stage of cluster galaxy evolution occurs; the metamorphosis into quiescent early-types observed by $z = 1$. An exploration of this desert holds great promise in unmasking this transition and providing a tighter framework for galaxy formation models to operate within.

With Pan-STARRS data, one may follow the cluster population to redshifts beyond those currently probed by the SDSS as the 4000Å spectral feature moves into the y-band filter. The Medium Deep Surveys (MDS), with deep y-band photometry, will be exploited to detect ∼1,000 clusters out to $z = 1.5$. Characterising the cluster y-band red sequence evolution in a sample of MDS clusters is the first step in undertaking a deep search of Pan-STARRS data. From a preliminary search through an early-release MDS field (9 square degrees), we identified 25 clusters with detectable $i - y$ red sequences. The Brightest Cluster Galaxies (BCG) of these clusters feature among a sample we constructed being spectroscopically targeted for preliminary investigations of the sequence evolution. Following completion of PS-1 operations, compilation of the most comprehensive optically-selected deep cluster catalogue will enable both direct and follow-up studies to build a picture of the evolving cluster membership and quantities such as the Dark Energy equation of state parameter w. Because near-IR luminosities are relatively insensitive to ongoing or recent star formation activity, access to *J*, *H* and *K*-band photometry could probe cluster stellar mass assembly while the ORCA detector searches to higher redshifts.

Whilst the total area covered by the MDS is some $50\times$ smaller than that proposed by the Dark Energy Survey, PS-1 will probe to deeper redshifts, and sooner, providing early precision-estimates of w through cluster counts. Moreover, while plans to add a Y-band filter to the existing DES bands will allow the detection of higher-redshift clusters, it will require an understanding of the red sequence evolution proposed here.

Unmasking cluster evolutionary history: refining E/SO ridgeline detections

By following the 4000Å spectral feature to redder wavelengths, one is able to track the evolution of the cluster population for as long as a detectable red sequence remains in place. With a fast-growing volume of IR data (e.g. *Herschel*, *Spitzer*), and evidence that

galaxy colour bi-modality is in place as early as $z = 3$ (Kodama et al., 2007; Doherty et al., 2010), logical extensions to the algorithm are modifications to search for more distant clusters. An ORCA adaptation has already identified clusters in SERVS[1] with near-IR *Spitzer* bands (Geach et al., 2010) out to $z = 1.2$, but work towards sampling the $z = 1.2 - 2$ regime will provide direct access to what is at present only a sparse sample of distant ETGs.

Star formation in Brightest Cluster Galaxies (BCGs)

With a large sample of BCG candidates, one may probe the role star formation plays in the cores of clusters. Excess blue light from short-lived main-sequence stars leads to departures from the red sequence in the BCG population, impacting cluster detection algorithms such as ORCA .

As an external member of the SDSS3:BOSS team, we will assist in probing the star formation and AGN activity of ~ 1500 cluster BCGs we selected from the Stripe 82 cluster catalogue in chapter 2 (PI: Leauthaud). This sample, the largest of its kind, will provide a comprehensive probe of central star formation out to $z = 0.4$. Although different feedback prescriptions make very distinct predictions on the amount of star formation in massive haloes (McCarthy et al., 2011), only observational data samples of this size can provide the resolving power to discriminate between them. Enhancing this dataset are spectra currently being taken of cluster BCGs in the Pan-STARRS MDS fields. Follow-up, high-resolution (or IFU) spectroscopy of selected targets would yield a low-risk insight into the complex behaviour of cluster cores.

Detecting ORCA-lensed "red nuggets"

With the gravitationally lensed amplification of faint high-redshift galaxies by ORCA clusters, near-IR photometric filters are well placed to identify "red nuggets" - compact Early Type Galaxies (ETGs) at $z = 1 - 2$ with evolved stellar populations and low levels of star formation. Accounting for the assembly and subsequent evolution of such systems into today's population of massive early-type ellipticals presents a major challenge to our understanding of galaxy evolution (see, for example Newman et al., 2010). With *Herschel* data, (Coppin et al., 2011) have already identified significant excess far-IR emission attributed to (potentially lensed) background sources within the virial radius of SDSS-

[1] http://www.cv.nrao.edu/∼mlacy/servs.html

selected ORCA clusters.

As co-I of a Cycle 8 *Spitzer* Snapshot Proposal (PI: Geach), we hope to secure 3.6μm and 4.5μm imaging of $\sim$400 ORCA SDSS Stripe 82 cluster lenses at $\mathrm{z} = 0.3 - 0.5$, aiding in the identification of at least 50 lensed ETGs. In the $\mathrm{z} = 1 - 2$ target redshift range, the remaining IRAC channels probe the rest-frame 1.6μm H^{-} spectral bump of early-type galaxies. Supporting these bands with existing optical (e.g, *r*-band) imaging allows a clean colour-colour separation between this high-redshift population and members of the lensing cluster itself. Dust-obscured high redshift galaxies can be filtered by 3.6μm flux, whilst detections of redder, dust-enshrouded AGN provide (gratis) a supplementary dataset. Follow-up spectroscopy required to characterise these galaxies would not require the prohibitively expensive integration times required of an equivalent, unlensed sample. For example, had the J1255-0 (Kriek et al., 2009) ETG been a lensed red nugget, amplified by $5\times$, one could have reduced the heroic $\sim$30hr slit integration time to 1.2hr for the same signal-to-noise spectrum. Dramatically reducing the required integration time for each object in this way facilitates rapid compilation of the first comprehensive "nugget" sample through which a clearer understanding of their properties and characteristics can be gained. While this work carries some level of risk (e.g. fewer lensed sources than anticipated), the potential number of high-redshift candidate ETGs is competitive with the *HST* CANDELS[2] project by trading depth and spectral coverage with area.

Cluster, filament & void cosmology: the Large-Scale Structure link

Evolution of the cluster mass function is a sensitive probe of structure growth and the distance-redshift relation. These in turn rely on the properties of dark energy (characterised by the equation of state parameter w), and the mass density of the universe Ω_{M}. Characterising the ORCA cluster mass function in a large survey volume such as Pan-STARRS, combined with priors from other experiments (e.g. Baryonic Acoustic Oscillation studies and the CMB) will lead to tighter cosmological parameter constraints.

The filament-detection approach laid out in chapter 2 will be extended to identify ORCA-based superstructures not only in a survey volume approximately $40\times$ that of the 2dFGRS, but also in an array of realistic mock surveys. Filament catalogues can place

[2] http://candels.ucolick.org

limits on the frequency of extreme ($L_{b_J} > 10^{13.5}\,h^{-2}\,\mathrm{L}_\odot$) structures in the real universe, and the ability of models to reproduce them in the same abundance. Quantitative comparisons of mock filament catalogues with differing cosmologies will allow an evaluation of filaments and superclusters as probes of cosmological parameters such as σ_8. By exploiting the large volumes probed with PS-1 3π data, the detection of new "superstructures" may place limits on the homogeneity scale required to conserve the cosmological principle.

Non-Gaussianities in the primordial density distribution will be most apparent from the statistics of objects characterising its extreme tail. In conjunction with void catalogues, the statistical analysis of large-volume filamentary catalogues could lead to constraints on the non-Gaussianity parameter f_{NL} independent of those from higher-order galaxy clustering and biasing in the halo power spectrum (see, e.g. Verde, 2010). These data could even hint at discordance in the inflationary model of structure formation.

7.3.2 Developing the ORCA detector

We identify here two key projects that will produce higher quality science through additional improvements to the algorithm:

Optical richness estimators and mass calibration

We are involved in a broad collaborative drive to enhance science from SDSS Stripe 82, transforming it into a valuable multi-epoch, multi-wavelength data legacy. For example, we assisted in defining the footprint for a deep, <0.8"-seeing CFHT imaging survey (PI: Kneib, close to completion at time of writing) that will cover many of the clusters we detected in chapter 4. The imaging quality of this survey is ideal for constraining the masses of ORCA clusters through weak lensing.

Connecting optical richness measures (e.g. $\mathrm{B_{gc}}$) to more direct mass measurements of a cluster sample is observationally the cheapest way to derive masses for clusters spanning wide-area optical surveys, and important in determining the selection limits of the cluster catalogue produced. However, the correlation between optical richness and cluster mass estimates (X-ray or otherwise) suffers from large scatter (see Gladders et al., 2007; Rozo et al., 2008, for examples). We aim to formulate an improved cluster

richness parameter that could be applied to the optical counterparts of, for example SZE-detected clusters. Preliminary simulation tests of optical mass proxies we described in chapter 4 will be connected to five inter-calibrated cluster mass estimators:

i Weak lensing studies of Stripe 82 ORCA clusters with deep CFHT i-band imaging (PI: Kneib).

ii X-ray cluster masses with proposed (PI: Glikman) deeper XMM coverage complementary to CFHT data.

iii Strong-lensing model cluster mass estimates (e.g. LENSTOOL, Jullo et al., 2007) using a visually-identified Stripe 82 arc catalogue generated in Durham.

iv Exploiting the $M^{\star} - M_{\rm halo}$ connection with near-IR stellar masses via proposed NEWFIRM (PI: Ross) imaging.

v Redshift-insensitive SZE mass estimates with stacked Planck data (via the Pan-STARRS MOU) and the ACT cluster detection campaign where possible.

Constraining the cluster selection function

In cluster science, the results from the weighting, stacking or binning of systems relies as much on what isn't observed as what is. In chapter 3, we use mock data to quantify the algorithm performance in terms of the completeness, purity and spurious detection rate. Although tests indicate a detection performance on a par with alternative algorithms, further work in this area is essential for the production of cluster-led science. To this end, work is underway to compare the ORCA cluster selection function to that of a complementary group finder (Liu et al., 2008) using PS-1 MDS data. With our experience in generating realistic mock surveys, a suite of 50 independent MDS reproductions will be derived from Pan-STARRS lightcone data, providing a deeper insight into the detection capability. Moreover, the comparison of observed cluster properties to those in mock catalogues will motivate the further development required (e.g. Hilbert and White, 2010; de la Torre et al., 2011) in simulating the observed group and cluster galaxy population.

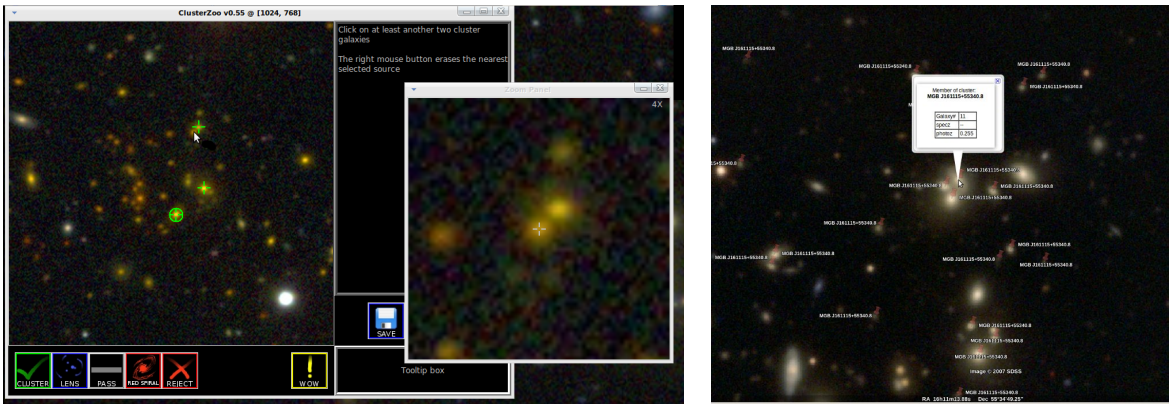

Figure 7.1: *Left:* Screenshot of the prototype *Cluster Zoo* utility. A non-expert used this application to identify distant galaxies gravitationally lensed by ORCA clusters selected from chapter 4. From a sample of 1100 clusters, 450 putative lenses were classified. *Right:* Exploring ORCA Pan-STARRS galaxy clusters in *Google-Sky*.

7.3.3 Distribution, interaction and exploration of cluster data

Often overlooked in the challenge of deriving maximal science from a dataset is the "front end" - how (and indeed what) data is made available. The SDSS CAS[3], for example, has been an indispensable resource for the astronomical community, resulting in SDSS legacy data being among the most published and cited Astronomical resources. Our plans for developing two particular aspects of access to ORCA data are discussed here:

Access to cluster data

Surveys such as the 2dFGRS and SDSS have shown that the true science potential of a project can be more easily reached if data is available in the public domain. *How then to support access to this data?*

Enabling science from large datasets firstly requires easy access to them. For Pan-STARRS, a collaboration cluster database would likely integrate with the Published Science Products Subsystem (PSPS, Leader: Jim Heasley). To progress with this, a SQL database could be developed, in much the same spirit as the Virgo Consortium Mil-

[3] http://casjobs.sdss.org

lennium database[4], capable of serving top-level cluster information, detailed cluster membership data and linkage to the significantly larger photometric parent catalogue. Throughout the development process we would supply both real and mock cluster data, assist in schema development and the soak-testing of a prototype database.

Citizen science

Our initial experience in community science came through development of `Cluster-Zoo`, a prototype application enabling public users to view and classify cluster imaging (see the *left* panel of Figure 7.1). In addition to gauging the completeness and spurious detection rate within cluster catalogues, considerable motivation comes from the identification of distant, lensed galaxies. Notoriously difficult to identify in an automated fashion, we foresee *"humanware"* leading software in efforts to detect lenses in wide imaging surveys. `Cluster-Zoo` assisted an undergraduate summer student in the compilation of a putative lensed galaxy catalogue from *irg* colour imaging of Stripe 82 `ORCA` clusters. With this single volunteer viewing a high-mass subset of the cluster catalogue, 450 putative arcs were identified. Some 7% of these were professionally evaluated as high probability lenses and a selection were spectroscopically followed up.

In related developments, we have produced a utility converting `ORCA` cluster data into "keyhole markup language" (.kml) files for interactive use in `Google-Sky`. In addition to denoting the position of cluster members on the sky-map, data is also included that describes basic properties of the cluster. Such utilities have broad outreach potential. `Google-Sky` data for Pan-STARRS cluster catalogues have been made available internally to consortium members. Both of these projects require further feature-development, testing and supporting reference material before they are made publicly available.

7.4 Concluding remarks

The key results and findings of our study into the detection of large-scale structure in large astronomical surveys are:

- With full 3D galaxy data, we show filamentary structure can be detected in spec-

[4] `http://galaxy-catalogue.dur.ac.uk:8080/Millennium`

troscopic surveys, but mock surveys are as yet unable to reproduce the largest observed structures. Larger surveys may be required to characterise these superstructures and attempt precision "filamentary cosmology".

- Attempting recovery of the same structures using photometric redshifts tends to homogenise the connected systems, erasing the lowest and highest luminosity detections. The algorithms used to calculate these redshifts must reduce their measurement uncertainties to at least $\delta_z = 0.002$ to recover a similar range of structure luminosities found in spectroscopic datasets.
- We develop an algorithm to detect galaxy clusters that makes no prior assumptions about their properties other than the similarity in colour of their members and an enhanced projected surface density. Tests on model data with the algorithm reveal a detection performance matched well against alternative algorithms.
- Application of this algorithm to a 270deg^2 survey with deep SDSS photometry produced a catalogue of over 4,000 clusters. Scientific exploitation of this survey is ongoing, but studies in work presented here show: cluster identifications to $z = 0.6$, agreement with the model red sequence slope evolution, evidence stellar age is not responsible for the sequence, and a well-defined colour-colour track of potential use in photometric cluster redshift estimation.
- Following additional development, we find the algorithm is capable of successfully detecting galaxy clusters in a variety of datasets. Our largest study, covering 7,600 square degrees, recovered 32,808 clusters and provides an estimated cluster yield for the PS-1 3π survey. Using the accurate 3D positions of these clusters, we are able to recover connected structures also identified with spectroscopic data in the 2dFGRS.
- By detecting clusters in *iJK* band-merged photometry, we demonstrate `ORCA` is able to track the 4000Å break beyond the optical regime.
- A frequently used model describing Voronoi cell areas in a random distribution fails to accurately predict the high density regime important for cluster detection. The alternative model presented in this work provides a better fit to this distribution and may in the future be used to improve Voronoi-based recovery of clustered data.

Scientific Acknowledgements

In addition to the section at the front of this thesis, included here are acknowledgements to organisations who made available their data, code or support during the course of this work:

- The Science and Technology Facilities Council for financial support with a PhD studentship.
- The Hatfield Trust for an MCR Research Award.
- The 2dF Galaxy Redshift Survey was made possible through the dedicated efforts of the staff of the AAO, both in creating the two-degree field instruments and in supporting it on the telescope.
- Calculations in parts of this thesis were performed on the ICC Cosmology Machine, which is part of the DiRAC Facility jointly funded by STFC, the Large Facilities Capital Fund of BIS, and Durham University.
- The Millennium Simulation databases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory.
- Funding for the SDSS, SDSS-II and SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is `http://www.sdss.org`. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the

Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

- The Pan-STARRS Project is being led by the University of Hawaii Institute for Astronomy, and exploits the unique combination of superb observing sites and technical and scientific expertise available in Hawaii. Funding for the development of the observing system has been provided by the United States Air Force Research Laboratory. The PS1 Surveys have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network, Incorporated, the National Central University of Taiwan, and the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the author(s), and do not necessarily reflect the views of the National Aeronautics and Space Administration.
- The UKIDSS project is defined in Lawrence et al. (2007). UKIDSS uses the UKIRT Wide Field Camera (WFCAM; Casali et al., 2007). The photometric system is described in Hewett et al. (2006), and the calibration is described in Hodgkin et al. (2009). The pipeline processing and science archive are described in Hambly et al. (2008).
- The task of manipulating large datasets was eased considerably with our use of TOPCAT (Taylor, 2005). We thank Mark Taylor for the continued development of this utility.
- We thank Vitalii Vanovschi for the development, availability and support of the `Parallel-Python` module (`http://www.parallelpython.com`)

Appendix A

The redshift-dependent 2PIGG purity $f(\mathrm{z})$

To account for the incorrect inclusion of interloper galaxies in 2PIGG galaxy groups, in §2.4 we randomly eject a random, redshift-dependent fraction of group galaxies back into the field. From Eke et al. (2004a), we derive the fraction, $f(\mathrm{z})$, of group galaxies that are genuine members of the parent halo. We use two quantities described in §4 of that paper:

(i) $< N_{\rm good}/N_{\rm spawn} >$: the mean fraction of interloper galaxies assigned to haloes with $N_{\rm spawn}$ galaxies.

(ii) $< q >$: the mean quality of the group-to-halo matches

In Eke et al. (2004a), the "quality" parameter is defined as:

$$q = \frac{N_{\rm good} - N_{\rm bad}}{N_{\rm spawn}}, \tag{A.1}$$

where $N_{\rm good}$ is the number of galaxies assigned to a group that belong in the halo, $N_{\rm bad}$ is the number of galaxies assigned to the group that were not from the halo, and $N_{\rm spawn}$ is the true number of galaxies residing in halo.

Our definition of $f(\mathrm{z})$ is therefore:

$$f(\mathrm{z}) = \frac{N_{\rm good}}{N_{\rm good} + N_{\rm bad}}. \tag{A.2}$$

Our estimation of the redshift variation in the quantities (i) and (ii) describe above comes from Figure 2 of Eke et al. (2004a) (reproduced here as Figure A.1):

$$\frac{N_{\rm bad}}{N_{\rm spawn}} = 0.2 + \mathrm{z}, \frac{N_{\rm good} - N_{\rm bad}}{N_{\rm spawn}} = 0.8 - 1.5\mathrm{z}. \tag{A.3}$$

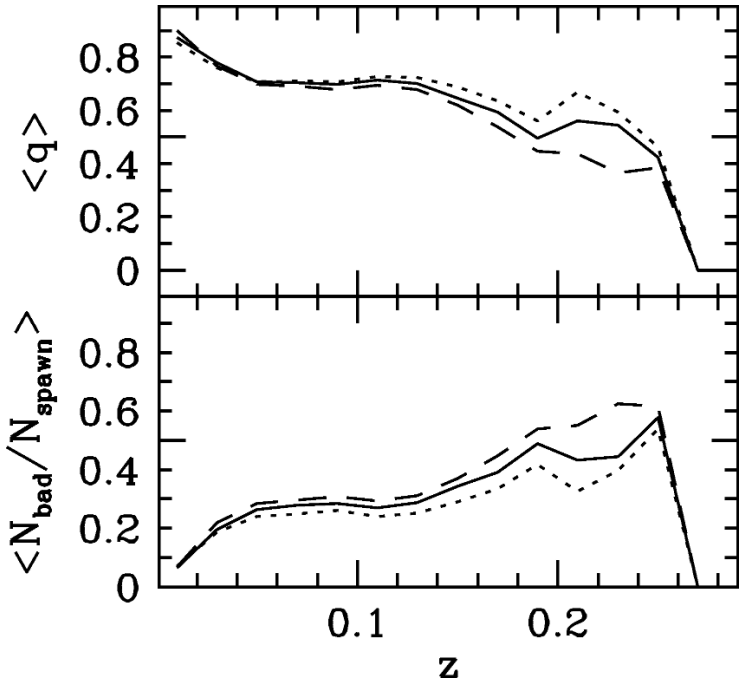

Figure A.1: (adapted from Eke et al., 2004a) The mean interloper fraction for haloes with N_{spawn} galaxies (*top*) and the quality parameter *bottom* The three lines indicate different aspect ratios R_{gal} of the linking cylinder. Our study uses R_{gal}=11 - the solid line.

From these definitions we derive $f(\mathrm{z})$:

$$\frac{N_{\mathrm{good}}}{N_{\mathrm{spawn}}} - 0.2 + \mathrm{z} = 0.8 - 1.5\mathrm{z}$$

$$\frac{N_{\mathrm{good}}}{N_{\mathrm{spawn}}} = 1 - 0.5\mathrm{z}$$

$$N_{\mathrm{good}} = (1 - 0.5\mathrm{z})N_{\mathrm{spawn}}$$

$$N_{\mathrm{good}} + N_{\mathrm{bad}} = N_{\mathrm{good}} - N_{\mathrm{good}} + 2N_{\mathrm{bad}} = (0.8 - 1.5\mathrm{z})N_{\mathrm{spawn}} + 2N_{\mathrm{bad}}$$

$$N_{\mathrm{good}} + N_{\mathrm{bad}} = (0.8 - 1.5\mathrm{z})N_{\mathrm{spawn}} + 2(0.2 + \mathrm{z})N_{\mathrm{spawn}}$$

$$N_{\mathrm{good}} + N_{\mathrm{bad}} = (1.2 + 0.5\mathrm{z})N_{\mathrm{spawn}}$$

$$\frac{N_{\mathrm{good}}}{N_{\mathrm{good}} + N_{\mathrm{bad}}} = \frac{1 - 0.5\mathrm{z}}{1.2 + 0.5\mathrm{z}} = \frac{2 - \mathrm{z}}{2.4 + \mathrm{z}} = f(\mathrm{z}). \tag{A.4}$$

Appendix B | *SDSS CAS Query*

This query was used to extract model magnitudes and their errors for galaxies in both the small 7deg^2 subset in chapter 3 and the full survey in chapter 4. A slightly modified version of this query was used to extract galaxies for the SDSS DR7 study in chapter 5.

```
SELECT      p.ra,p.dec, p.objID

            p.modelMag_u as model_u,p.modelMag_g as model_g,
            p.modelMag_r as model_r, p.modelMag_i as model_i,
            p.modelMag_z as model_z,

            p.modelMagErr_u as model_u_err,p.modelMagErr_g as
            model_g_err, p.modelMagErr_r as model_r_err,
            p.modelMagErr_i as model_i_err, p.modelMagErr_z as
            model_z_err,

            p.extinction_u as ext_u, p.extinction_g as ext_g,
            p.extinction_r as ext_r, p.extinction_i as ext_i,
            p.extinction_z as ext_z

FROM        Stripe82..Galaxy as p

WHERE       (flags & (dbo.fPhotoFlags('BINNED1') |
            dbo.fPhotoFlags('BINNED2') |
            dbo.fPhotoFlags('BINNED4'))) > 0
            AND (flags & (dbo.fPhotoFlags('BLENDED') |
            dbo.fPhotoFlags ('NODEBLEND') |
```

```
dbo.fPhotoFlags('CHILD'))) != dbo.fPhotoFlags('
    BLENDED')
AND (flags & (dbo.fPhotoFlags('EDGE') | dbo.
    fPhotoFlags
('SATURATED'))) = 0
AND (p.psfMag_r-p.modelMag_r > 0.05)
AND ((p.run = 106 OR p.run = 206)
AND (p.modelMag_r > 14.0)
AND ((p.modelMag_r-p.extinction_r) <= 24.0))
```

The conditions listed at the end of this SQL query ensure the objects returned are genuine galaxies with high quality photometry. In particular, the image processing flags from the SDSS photometric pipeline are used to ensure stellar contamination and low-significance detections are filtered from the catalogue. The flags we make use of are briefly summarised below:

- `BINNEDn`: (n=1,2,4) An object detected with greater than 5-σ significance in a binned n$\times$n image.
- `BLENDED`: Object has more than one peak, and so may be a composite source.
- `NODEBLEND`: Object has more than one peak, but was not deblended (generally due to its proximity to an edge)
- `CHILD`: The object is the result of an attempt to deblend the parent. We included these objects where it is not also a candidate for deblending.
- `EDGE`: Where an object is too close to the edge of the frame, it will be assigned this flag. These objects should be re-detected in overlaps, and are therefore rejected in our query.
- `SATURATED`: The object contains saturated pixels. We reject objects that have been assigned this status.

Appendix C | *Cluster data*

This appendix contains HDF5 Stripe 82 catalogue information and top-level data from ORCA catalogues. Readers are referred to §3.4.1 for a description of the table-data columns.

C.1 Stripe 82 hierarchical data

This section describes the schema of the Stripe 82 Hierarchical Data Format (HDF, version 5) cluster catalogue; this format permits access to more detailed cluster data.

Parameter	Description	Ref
ID	Cluster identification number	
ra, dec	Cluster RA, dec (°, J2000)	§3.4.2, p.99
ra_bcg, dec_bcg	BCG RA, dec (°, J2000)	
ngal	Number of cluster galaxies	
redshift	Cluster redshift	§3.4.2, p.99
redshift_code	Source of redshifts	§3.4.2, p.99
theta80	Cluster angular radius θ_{80} (°)	§3.4.5, p.101
concentration	θ_{80}/θ_{20} concentration parameter	§3.4.5, p.101
Agc	Preferred richness metric $A_{\rm gc[1,i]}$	§4.4.2, p.143
/. ./Richness		
../Class	Richness class R1-R4	§4.4.2, p.144
../R500/	Measurements within $0.5h^{-1}$Mpc	§4.4.2, p.142
../Theta80/	Measurements within θ_{80}	§4.4.2, p.142
../../RedSequence/	Galaxies within selection filter	§4.4.2, p.142
../../MagLim/	Galaxies down to flux limit m_l	§4.4.2, p.142
../../../../Nb	Background galaxy counts	§4.4.2, p.142
../../../../Nnet	Net galaxy counts	§4.4.2, p.142
../../../../Agc	Angular correlation function amplitude	§4.4.2, p.143
../../../../Bgc	$B_{\rm gc}$ richness statistic	§4.4.2, p.143
/. ./Galaxies/GalaxyNNN		
../objID, DR7_objID	Unique SDSS Stripe 82/DR7 identifier	
../ra, dec	RA, dec (°, J2000)	
../specz, photoz	Spectroscopic and photometric redshifts	
../specz_source	Spectroscopic redshift source	
../photoz_source	Photometric redshift source	
../u, g, r, i, z	SDSS model magnitudes	

Table C.1: Information and references for the schema in Figure C.1.

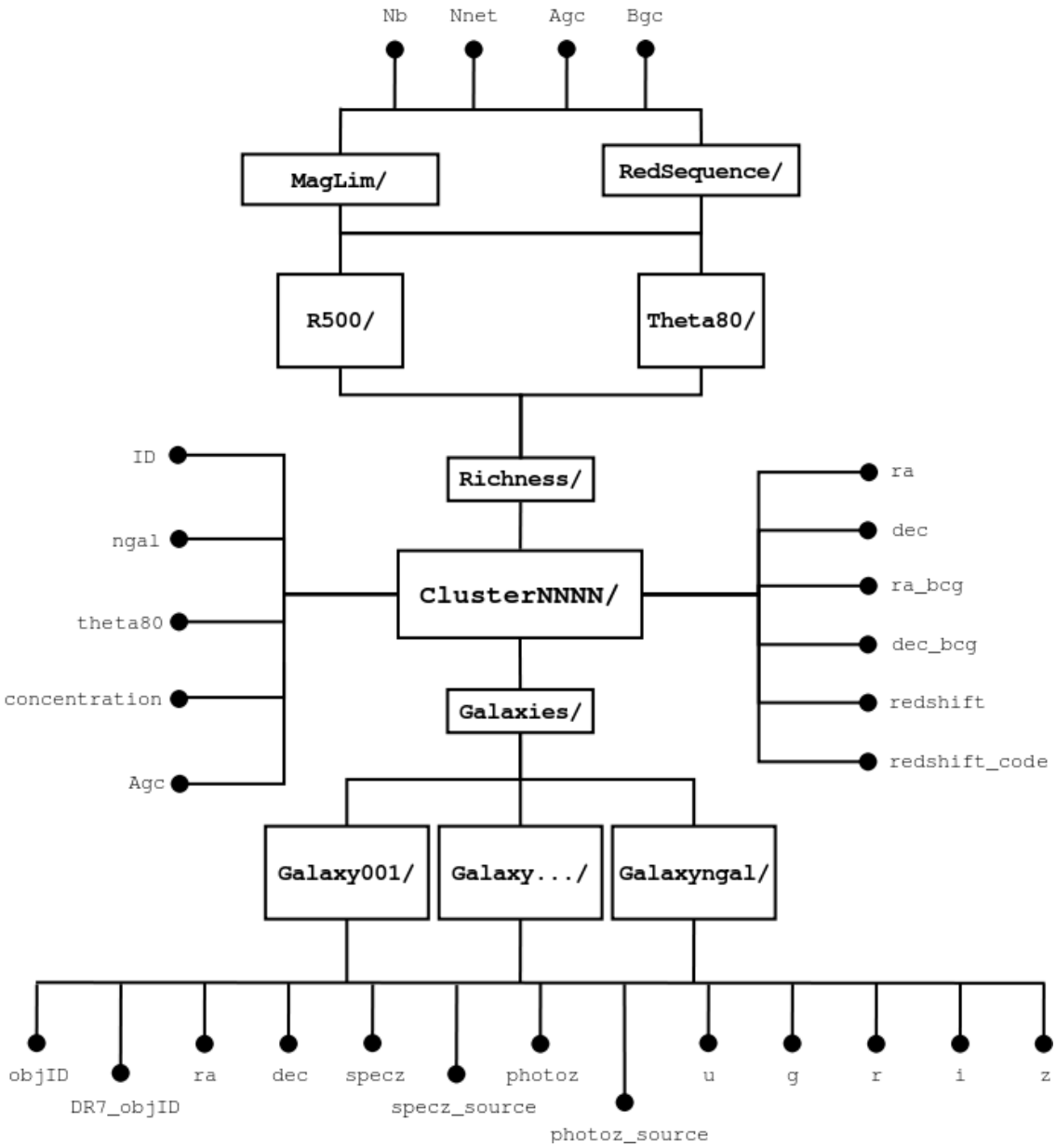

Figure C.1: HDF5 schema for the ORCA Stripe 82 cluster catalogue. The catalogue comprises a list of clusters (`Cluster0000`-`Cluster4097`). In addition to the "top level" data similar to those available in the tables of this appendix, each cluster also has daughter tables containing richness and member galaxy data.

C.2 Small Stripe 82 region

Table C.2 – Stripe 82 7deg^2 sample

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MGB J233730+00430.4	354.37650	0.72415	0.258	c0s0w0q0d0b0p7h7	7	2322	0.015	0.0002	2.625
MGB J234351-01110.9	355.96333	-1.19785	0.292	c0s0w0q0d0b0p8h8	8	931	0.029	0.0003	2.750
MGB J234251+01070.3	355.71444	1.12172	0.349	c0s0w0q0d0b0p5h7	7	1666	0.038	0.0002	2.091
MGB J234105+00180.3	355.27138	0.30514	0.259	c0s0w1q0d0b0p58h55	66	3744	0.032	0.0007	2.545
MGB J233740+00160.2	354.41537	0.26931	0.257	c0s4w2q0d0b0p167h160	174	4849	0.030	0.0012	3.000
MGB J234548-01070.6	356.45011	-1.12640	0.272	c0s0w0q0d0b0p24h28	28	3690	0.033	0.0004	2.385
MGB J234447-00050.2	356.19500	-0.08674	0.269	c0s1w0q0d0b0p8h4	8	3350	0.027	0.0001	2.571
MGB J233735-00400.1	354.39529	-0.66782	0.284	c0s1w1q0d0b0p43h45	47	3396	0.032	0.0006	2.500
MGB J234720+00290.6	356.83294	0.49251	0.265	c0s0w0q0d0b0p8h11	11	2185	0.030	0.0002	1.467
MGB J234351-01090.1	355.96288	-1.15232	0.275	c0s0w0q0d0b0p4h0	5	2159	0.028	0.0001	1.500
MGB J234053+00050.7	355.22174	0.09534	0.171	c0s0w0q0d0b0p20h21	21	5934	0.047	0.0003	2.500
MGB J233903+00430.7	354.76447	0.72874	0.184	c0s1w0q0d0b0p12h10	12	3693	0.031	0.0003	3.833
MGB J234105+00040.5	355.26926	0.07511	0.183	c0s1w0q0d0b0p13h5	14	2090	0.035	0.0004	1.889
MGB J234608+00580.9	356.53140	0.98227	0.093	c0s1w0q0d0b0p4h4	5	-12360	0.047	0.0002	1.706
MGB J234736-00360.5	356.89901	-0.60895	0.187	c0s0w0q0d0b0p5h2	5	1701	0.015	0.0001	2.000
MGB J233815+00060.5	354.56325	0.10819	0.248	c0s0w0q0d0b0p5h0	5	6304	0.043	0.0001	2.222

Continued overleaf...

Table C.2 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MGB J234312-00510.6	355.80111	-0.85918	0.267	c0s1w0q0d0b0p7h4	7	1938	0.035	0.0002	3.000
MGB J233944+00400.8	354.93332	0.67918	0.413	c0s0w0q0d0b0p7h7	7	1518	0.026	0.0005	1.692
MGB J233635-00470.1	354.14686	-0.78419	0.442	c0s0w0q0d0b0p6h5	7	2337	0.049	0.0004	1.500
MGB J234310+00590.8	355.79037	0.99674	0.453	c0s0w0q0d0b0p5h5	5	1726	0.014	0.0003	5.000
MGB J234341+00180.3	355.91903	0.30526	0.240	c0s2w0q0d0b0p130h130	133	2654	0.039	0.0012	3.200
MGB J233650+00140.7	354.20660	0.24425	0.196	c0s0w0q0d0b0p9h9	9	5710	0.036	0.0003	2.357
MGB J234647+00570.6	356.69566	0.95975	0.222	c0s0w0q0d0b0p5h5	5	2355	0.048	0.0002	8.333
MGB J234603+01080.4	356.51454	1.14007	0.394	c0s0w0q0d0b0p4h5	5	802	0.029	0.0001	4.333
MGB J234134+00420.8	355.39320	0.71322	0.475	c0s0w0q0d0b0p3h1	5	706	0.027	0.0001	2.000
MGB J234607+00420.6	356.52779	0.71050	0.460	c0s0w0q0d0b0p4h5	5	867	0.037	0.0002	1.667
MGB J233840-00190.1	354.66559	-0.31811	0.292	c0s1w0q0d0b0p4h4	5	999	0.023	0.0001	2.000
MGB J233902+00100.0	354.75721	0.16736	0.263	c0s1w0q0d0b0p7h6	7	1367	0.021	0.0001	1.444
MGB J234017+00530.1	355.07215	0.88444	0.330	c0s0w0q0d0b0p7h5	9	4501	0.022	0.0002	2.286
MGB J234625+00410.7	356.60551	0.69514	0.400	c0s0w0q0d0b0p8h7	8	2590	0.050	0.0002	4.250
MGB J233731-00070.0	354.37821	-0.13251	0.452	c0s1w0q0d0b0p9h8	9	1446	0.028	0.0002	5.000
MGB J233733-00010.0	354.38683	-0.01676	0.431	c0s0w0q0d0b0p14h15	18	1127	0.043	0.0005	2.182
MGB J234233-00150.8	355.63876	-0.26290	0.257	c0s0w0q0d0b0p12h13	13	3447	0.039	0.0005	5.500
MGB J233835+00130.7	354.64429	0.22839	0.232	c0s0w0q0d0b0p6h6	6	2165	0.039	0.0002	2.111

Continued overleaf...

Table C.2 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MGB J234100-00070.8	355.25003	-0.13003	0.440	c0s0w0q0d0b0p4h5	5	1967	0.021	0.0005	1.923
MGB J234734+00030.7	356.88994	0.06120	0.435	c0s0w0q0d0b0p6h6	7	1875	0.032	0.0003	2.167
MGB J234729-00080.4	356.86972	-0.13958	0.227	c0s0w0q0d0b0p52h53	53	6430	0.037	0.0008	2.600
MGB J233942+00450.3	354.92295	0.75510	0.233	c0s0w0q0d0b0p5h5	5	4313	0.043	0.0002	3.000
MGB J233926+00310.2	354.85816	0.52027	0.402	c0s0w0q0d0b0p5h7	7	2726	0.043	0.0003	1.455
MGB J234523+00260.8	356.34654	0.44640	0.421	c0s0w0q0d0b0p13h13	13	2185	0.030	0.0007	2.500
MGB J234453+00410.4	356.21899	0.69040	0.349	c0s0w0q0d0b0p5h5	5	-839	0.043	0.0002	1.636
MGB J233939+01010.4	354.91346	1.02381	0.351	c0s0w0q0d0b0p5h4	5	410	0.008	0.0001	2.667
MGB J233829+00200.0	354.61883	0.34948	0.443	c0s0w0q0d0b0p6h5	6	654	0.042	0.0001	1.400
MGB J233734+00060.2	354.39220	0.10261	0.536	c0s0w0q0d0b0p4h5	5	524	0.026	0.0003	1.167
MGB J233629-01060.0	354.11913	-1.11592	0.198	c0s1w0q0d0b0p4h4	5	-1494	0.025	0.0001	2.143
MGB J234424+00270.0	356.09845	0.45041	0.256	c0s0w0q0d0b0p5h4	5	1913	0.010	0.0002	3.167
MGB J234023+00560.1	355.09672	0.93484	0.358	c0s1w0q0d0b0p26h23	31	178	0.033	0.0016	1.750
MGB J234553+01100.6	356.47176	1.17619	0.338	c0s0w0q0d0b0p5h5	5	2719	0.038	0.0008	1.714
MGB J233902-00500.9	354.75878	-0.84845	0.447	c0s1w0q0d0b0p6h5	6	2706	0.022	0.0004	1.786
MGB J234119-00350.4	355.33113	-0.58965	0.418	c0s0w0q0d0b0p6h4	6	1097	0.031	0.0005	3.167
MGB J234400-00300.3	355.99952	-0.50461	0.181	c0s1w0q0d0b0p5h4	6	4477	0.025	0.0001	1.750
MGB J233826-01140.7	354.60916	-1.24432	0.186	c0s0w0q0d0b0p5h6	6	2502	0.042	0.0003	3.714

Continued overleaf...

Table C.2 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MGB J234105-00000.1	355.26969	-0.00103	0.189	c0s1w0q0d0b0p7h5	7	3588	0.033	0.0002	1.643
MGB J233855+01070.1	354.72952	1.11882	0.117	c0s0w0q0d0b0p6h6	6	-4465	0.011	0.0001	1.833
MGB J234440+00310.6	356.16517	0.52732	0.520	c0s0w0q0d0b0p6h3	6	1090	0.019	0.0002	1.571
MGB J234507-01120.2	356.27820	-1.20353	0.520	c0s0w0q0d0b0p4h0	6	918	0.051	0.0005	2.182
MGB J234605-00100.8	356.52145	-0.17945	0.232	c0s0w1q0d0b0p23h22	23	2599	0.025	0.0005	3.250
MGB J234626+00440.3	356.60782	0.73829	0.245	c0s1w0q0d0b0p50h50	53	2947	0.031	0.0008	2.583
MGB J234017-00030.9	355.06912	-0.06455	0.245	c0s0w0q0d0b0p6h2	6	2879	0.047	0.0001	1.700
MGB J234534-00150.1	356.39036	-0.25245	0.250	c0s0w0q0d0b0p5h5	5	2476	0.025	0.0001	1.429
MGB J233638-00360.3	354.15925	-0.60556	0.420	c0s0w0q0d0b0p4h5	5	802	0.010	0.0002	4.750
MGB J233812+00140.4	354.54942	0.24079	0.463	c0s0w0q0d0b0p5h5	5	1778	0.045	0.0002	3.143
MGB J234733-00100.0	356.88907	-0.18285	0.533	c0s0w0q0d0b0p7h0	7	2037	0.030	0.0003	2.571
MGB J234251-00140.8	355.71424	-0.24696	0.504	c0s0w0q0d0b0p5h5	6	1127	0.032	0.0001	1.143
MGB J234153-00220.8	355.47120	-0.38032	0.318	c0s1w0q0d0b0p7h8	9	3739	0.028	0.0003	4.250
MGB J233742-00150.9	354.42362	-0.26453	0.291	c0s1w0q0d0b0p5h5	6	1868	0.037	0.0002	3.000
MGB J234509-01090.8	356.28767	-1.16266	0.375	c0s0w0q0d0b0p7h8	9	2342	0.038	0.0003	2.231
MGB J234618+01020.0	356.57499	1.03369	0.397	c0s1w0q0d0b0p4h4	5	1218	0.018	0.0001	2.200
MGB J234426+00550.5	356.10916	0.92463	0.350	c0s1w0q0d0b0p5h3	5	1064	0.019	0.0002	2.667
MGB J234259-00100.6	355.74446	-0.17649	0.270	c0s0w0q0d0b0p5h5	5	557	0.027	0.0001	10.000

Continued overleaf...

Table C.2 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MGB J234517-00410.0	356.31923	-0.69994	0.573	c0s0w0q0d0b0p4h5	5	642	0.039	0.0004	4.167
MGB J234022-00040.0	355.09272	-0.08280	0.271	c0s0w0q0d0b0p18h18	18	2985	0.042	0.0005	1.500
MGB J234233-00460.1	355.63596	-0.76864	0.200	c0s0w0q0d0b0p4h5	5	1159	0.015	0.0001	4.333
MGB J234443+01070.9	356.18091	1.13198	0.356	c0s1w0q0d0b0p11h6	11	1860	0.042	0.0003	2.167
MGB J233934+00400.5	354.89040	0.67536	0.464	c0s0w1q0d0b0p6h6	6	1701	0.035	0.0002	1.750
MGB J234342+00460.3	355.92471	0.77232	0.468	c0s0w0q0d0b0p5h5	5	1932	0.018	0.0002	1.909
MGB J234526+00350.1	356.35697	0.58559	0.451	c0s0w0q0d0b0p4h5	5	2234	0.024	0.0003	2.625
MGB J233959-00250.9	354.99741	-0.43172	0.309	c0s0w0q0d0b0p8h7	8	2062	0.007	0.0005	3.000
MGB J233806+00080.4	354.52520	0.14030	0.155	c0s0w0q0d0b0p7h8	8	11192	0.025	0.0003	4.250
MGB J234140+01070.4	355.41634	1.12351	0.234	c0s1w0q0d0b0p21h20	22	3317	0.034	0.0006	2.300
MGB J233906-01130.2	354.77497	-1.22032	0.250	c0s0w0q0d0b0p10h11	11	1251	0.026	0.0005	3.556
MGB J234619+00430.4	356.57952	0.72407	0.113	c0s0w0q0d0b0p5h5	5	40259	0.021	0.0002	4.200
MGB J234358+00050.5	355.99021	0.09195	0.257	c0s0w0q0d0b0p9h9	9	1843	0.042	0.0004	3.667
MGB J234113-00000.4	355.30349	-0.00597	0.166	c0s0w0q0d0b0p6h2	6	15338	0.018	0.0003	1.692
MGB J233853+00310.1	354.71903	0.51834	0.491	c0s0w0q0d0b0p6h5	7	1792	0.039	0.0003	2.000
MGB J234213-01050.9	355.55221	-1.09907	0.351	c0s1w0q0d0b0p9h7	10	1977	0.028	0.0002	2.000
MGB J234725+00190.7	356.85322	0.32867	0.201	c0s0w1q0d0b0p14h14	14	1441	0.037	0.0004	2.545
MGB J234354+00570.0	355.97348	0.95027	0.347	c0s0w0q0d0b0p8h6	8	1390	0.016	0.0001	1.667

Continued overleaf...

Table C.2 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MGB J234218+00410.6	355.57459	0.69267	0.462	c0s0w0q0d0b0p5h4	6	1112	0.037	0.0001	1.667
MGB J234129-00250.1	355.36915	-0.41764	0.414	c0s0w0q0d0b0p8h7	8	1048	0.020	0.0003	2.167
MGB J233924-00170.0	354.85111	-0.28359	0.327	c0s1w0q0d0b0p5h4	5	3209	0.045	0.0003	1.333
MGB J234607+00040.6	356.52835	0.07686	0.408	c0s1w0q0d0b0p11h12	14	0	0.041	0.0009	2.062
MGB J233817+00190.0	354.56897	0.33309	0.208	c0s0w0q0d0b0p8h6	8	7445	0.038	0.0003	3.667
MGB J233815-00350.8	354.56316	-0.59662	0.327	c0s0w0q0d0b0p7h6	8	920	0.017	0.0002	2.571
MGB J233825-00100.1	354.60351	-0.16762	0.245	c0s0w0q0d0b0p15h16	16	1690	0.048	0.0005	4.250
MGB J234155+00400.4	355.48053	0.67314	0.401	c0s0w0q0d0b0p5h3	5	1615	0.024	0.0001	7.000
MGB J234130-01100.8	355.37491	-1.17994	0.405	c0s0w0q0d0b0p17h18	19	2257	0.039	0.0003	3.571

Table C.2: Data on the 97 ORCA clusters detected from the 7deg^2 sample investigated in Chapter 3.

C.3 Pan-STARRS MD-8 Data

Table C.3 – Pan-STARRS-1 MD-8 catalogue

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MD08 J161257+55480.4	243.23682	55.80725	0.069	c0s0w0q0d0p4h0	6	27579	0.025	0.0002	1.800
MD08 J161315+56010.1	243.31204	56.01843	0.445	c0s0w0q0d0p5h0	6	5113	0.011	0.0002	1.692
MD08 J161828+54500.2	244.61664	54.83655	0.223	c0s0w0q0d0p4h0	5	14970	0.023	0.0002	5.333
MD08 J161653+55080.2	244.22075	55.13708	0.364	c0s0w0q0d0p10h0	10	3281	0.027	0.0002	3.333
MD08 J161541+55070.8	243.91959	55.13017	0.295	c0s0w0q0d0p5h0	8	2636	0.020	0.0003	6.200
MD08 J161615+54400.9	244.06218	54.68171	0.291	c0s0w0q0d0p8h0	8	7073	0.057	0.0002	2.750
MD08 J161350+56160.1	243.45821	56.26844	0.132	c0s0w0q0d0p6h0	7	26896	0.042	0.0002	6.000
MD08 J161408+53250.9	243.53410	53.43200	0.229	c0s0w0q0d0p5h0	6	-5149	0.033	0.0006	2.833
MD08 J160951+55570.5	242.46141	55.95786	0.133	c0s0w0q0d0p12h0	13	-236	0.036	0.0006	4.000
MD08 J160417+56090.1	241.07270	56.15192	0.139	c0s0w0q0d0p6h0	6	1051	0.026	0.0002	1.700
MD08 J161438+54260.7	243.65648	54.44518	0.210	c0s0w0q0d0p27h0	33	7931	0.037	0.0007	1.917
MD08 J160506+55370.3	241.27547	55.62210	0.250	c0s0w0q0d0p39h0	51	5485	0.040	0.0013	4.000
MD08 J160112+54370.4	240.30130	54.62275	0.240	c0s0w0q0d0p6h0	7	17357	0.037	0.0005	2.059
MD08 J161244+55490.8	243.18522	55.82929	0.277	c0s0w0q0d0p13h0	13	4541	0.041	0.0005	3.000
MD08 J161858+54520.3	244.74333	54.87143	0.213	c0s1w0q0d0p8h0	10	-	0.036	0.0004	3.286
MD08 J161928+54130.0	244.86676	54.21721	0.272	c0s0w0q0d0p9h0	10	6812	0.042	0.0004	4.000

Continued overleaf...

Table C.3 – continued

Name	RA	dec	cluster_z	cz_type	N_{gal}	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MD08 J161644+56030.5	244.18257	56.05863	0.321	c0s0w0q0d0p8h0	9	5601	0.041	0.0003	4.000
MD08 J161707+54590.9	244.27836	54.99821	0.221	c0s0w0q0d0p6h0	6	24165	0.009	0.0003	1.714
MD08 J161115+55340.8	242.81218	55.58053	0.208	c0s0w0q0d0p48h0	53	4361	0.035	0.0008	3.375
MD08 J160804+54080.7	242.01732	54.14499	0.218	c0s1w0q0d0p8h0	10	228	0.039	0.0003	3.750
MD08 J161639+55400.2	244.16153	55.66945	0.227	c0s0w0q0d0p19h0	23	6148	0.037	0.0005	3.667
MD08 J161031+55480.5	242.62892	55.80843	0.276	c0s0w0q0d0p5h0	6	2465	0.028	0.0001	2.200
MD08 J161352+54470.6	243.46696	54.79349	0.314	c0s0w0q0d0p9h0	10	7160	0.017	0.0002	4.200
MD08 J161924+54000.1	244.84916	54.00241	0.269	c0s0w0q0d0p8h0	9	3044	0.034	0.0006	5.200
MD08 J161739+54120.3	244.41288	54.20571	0.329	c0s0w0q0d0p10h0	10	2918	0.014	0.0001	2.000
MD08 J161916+55130.5	244.81542	55.22571	0.323	c0s0w0q0d0p9h0	11	5445	0.018	0.0004	2.400
MD08 J160605+54280.7	241.51975	54.47836	0.678	c0s0w0q0d0p4h0	7	1276	0.023	0.0002	5.000
MD08 J161412+54440.1	243.55167	54.73510	0.302	c0s0w0q0d0p45h0	55	4788	0.017	0.0008	2.357
MD08 J161009+53460.9	242.53756	53.78221	0.453	c0s0w0q0d0p3h0	5	2202	0.010	0.0001	2.000
MD08 J160324+55060.1	240.85041	55.10155	0.268	c0s1w0q0d0p7h0	9	6803	0.015	0.0003	2.071
MD08 J160746+55060.9	241.94017	55.11505	0.205	c0s0w0q0d0p1h0	5	2165	0.017	0.0002	2.000
MD08 J160114+54400.0	240.30809	54.68275	0.312	c0s0w0q0d0p2h0	7	4024	0.043	0.0003	2.545
MD08 J160946+53490.9	242.44133	53.83130	0.238	c0s1w0q0d0p8h0	8	3649	0.012	0.0002	2.500
MD08 J160033+54400.6	240.13848	54.67596	0.166	c0s0w0q0d0p2h0	6	-7910	0.022	0.0008	1.692

Continued overleaf...

Table C.3 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MD08 J161935+53530.9	244.89774	53.89901	-	-	23	0	0.030	0.0016	1.667
MD08 J160635+54350.2	241.64714	54.58720	0.242	c0s1w0q0d0p7h0	10	7511	0.014	0.0002	3.000
MD08 J161032+55480.5	242.63218	55.80786	0.232	c0s0w0q0d0p17h0	22	5763	0.033	0.0006	2.556
MD08 J161254+55370.5	243.22313	55.62506	0.205	c0s0w0q0d0p5h0	5	435	0.009	0.0002	3.750
MD08 J161831+54170.5	244.62935	54.29220	0.253	c0s0w0q0d0p10h0	16	4059	0.039	0.0005	2.889
MD08 J160928+53470.9	242.36519	53.79848	0.062	c0s1w0q0d0p3h0	6	19290	0.021	0.0002	2.111
MD08 J161602+54440.3	244.00734	54.73864	0.228	c0s0w0q0d0p7h0	8	6316	0.017	0.0002	1.875
MD08 J161433+54460.7	243.63718	54.77772	0.322	c0s0w0q0d0p7h0	7	5513	0.041	0.0006	3.571
MD08 J160752+55320.0	241.96749	55.53406	0.251	c0s0w0q0d0p20h0	21	2138	0.044	0.0005	3.500
MD08 J160944+55170.7	242.43338	55.29510	0.261	c0s0w0q0d0p4h0	5	12206	0.033	0.0002	3.000
MD08 J160207+55360.3	240.52961	55.60456	0.153	c0s0w0q0d0p6h0	12	34603	0.046	0.0005	2.083
MD08 J160129+54510.4	240.37066	54.85700	0.266	c0s0w0q0d0p4h0	6	2444	0.033	0.0007	1.476
MD08 J161356+54270.2	243.48165	54.45335	0.287	c0s0w0q0d0p12h0	13	3577	0.044	0.0005	2.714
MD08 J161105+55470.7	242.77047	55.79477	0.244	c0s0w0q0d0p10h0	10	1420	0.029	0.0003	1.667
MD08 J160612+55050.1	241.55044	55.08420	0.221	c0s0w0q0d0p5h0	5	5744	0.024	0.0001	2.167
MD08 J160632+54560.8	241.63492	54.94605	0.225	c0s0w0q0d0p8h0	10	9392	0.030	0.0004	1.714
MD08 J160551+54460.7	241.46341	54.77830	0.242	c0s1w0q0d0p21h0	23	7109	0.029	0.0013	2.909
MD08 J160352+55500.3	240.96533	55.83811	0.203	c0s0w0q0d0p7h0	8	8249	0.009	0.0003	6.000

Continued overleaf...

Table C.3 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MD08 J161338+54180.9	243.40950	54.31540	0.387	c0s0w0q0d0p1h0	5	4942	0.014	0.0002	1.909
MD08 J160445+54500.7	241.18846	54.84532	0.232	c0s0w0q0d0p6h0	7	1342	0.016	0.0003	4.000
MD08 J161355+54470.9	243.47795	54.79885	0.385	c0s0w0q0d0p3h0	5	11740	0.016	0.0003	7.250
MD08 J161224+53400.9	243.10045	53.68101	0.226	c0s0w0q0d0p12h0	12	2191	0.043	0.0006	3.000
MD08 J160859+54180.9	242.24460	54.31420	0.263	c0s0w0q0d0p2h0	5	3087	0.036	0.0001	1.200
MD08 J161705+54500.4	244.26958	54.83992	0.422	c0s0w0q0d0p9h0	17	3078	0.032	0.0004	3.778
MD08 J161727+54330.4	244.36206	54.55736	0.266	c0s0w0q0d0p2h0	5	9928	0.022	0.0001	1.625
MD08 J161324+53420.8	243.34883	53.71276	0.275	c0s0w0q0d0p5h0	6	2954	0.011	0.0002	2.625
MD08 J162136+54580.7	245.40146	54.97905	0.229	c0s0w0q0d0p10h0	10	7990	0.025	0.0009	2.357
MD08 J161050+56170.2	242.70743	56.28670	0.250	c0s0w0q0d0p4h0	6	13727	0.034	0.0005	1.933
MD08 J161317+56010.2	243.31987	56.01990	0.471	c0s0w0q0d0p6h0	7	1027	0.019	0.0001	2.000
MD08 J161247+56070.6	243.19704	56.12601	0.376	c0s0w0q0d0p9h0	9	1080	0.030	0.0002	2.750
MD08 J160721+53540.1	241.83779	53.90097	0.117	c0s0w0q0d0p4h0	5	2540	0.038	0.0002	2.000
MD08 J160431+55130.0	241.13039	55.21671	0.336	c0s0w0q0d0p8h0	9	3801	0.024	0.0002	3.000
MD08 J161557+55100.2	243.98613	55.17028	0.359	c0s0w0q0d0p3h0	5	4482	0.013	0.0003	5.500
MD08 J161040+55260.3	242.66724	55.43817	0.231	c0s0w0q0d0p5h0	5	3980	0.009	0.0001	2.333
MD08 J161225+54010.4	243.10400	54.02295	0.338	c0s0w0q0d0p3h0	6	4461	0.020	0.0002	2.625
MD08 J161431+55090.0	243.62753	55.15060	0.181	c0s0w0q0d0p4h0	5	2064	0.032	0.0002	7.500

Continued overleaf...

Table C.3 – continued

Name	RA	dec	cluster_z	cz_type	N_{gal}	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MD08 J161226+55540.6	243.10892	55.91033	0.245	c0s0w0q0d0p22h0	23	4221	0.041	0.0009	4.600
MD08 J160437+55570.5	241.15293	55.95753	0.367	c0s1w0q0d0p8h0	14	-	0.020	0.0004	5.000
MD08 J160635+53530.8	241.64759	53.89599	0.348	c0s1w0q0d0p8h0	12	2458	0.015	0.0002	2.571
MD08 J160236+55100.9	240.64856	55.18145	0.243	c0s0w0q0d0p2h0	5	1531	0.016	0.0001	1.429
MD08 J161040+54060.0	242.66650	54.10004	0.330	c0s0w0q0d0p21h0	24	4088	0.016	0.0004	2.154
MD08 J160203+55230.5	240.51335	55.39166	0.228	c0s1w0q0d0p6h0	9	-	0.029	0.0003	2.667
MD08 J161324+53420.7	243.34832	53.71096	0.275	c0s0w0q0d0p5h0	6	13633	0.006	0.0002	2.750
MD08 J160342+56110.3	240.92345	56.18752	-	-	6	0	0.016	0.0011	4.333
MD08 J161257+54080.8	243.23839	54.14692	0.175	c0s0w0q0d0p5h0	6	2905	0.017	0.0001	2.167
MD08 J161925+55140.3	244.85384	55.23756	0.587	c0s0w0q0d0p4h0	5	1299	0.016	0.0001	2.286
MD08 J160412+55420.1	241.04837	55.70239	0.290	c0s0w0q0d0p7h0	9	1644	0.014	0.0002	3.400
MD08 J160231+54090.0	240.62803	54.15034	0.368	c0s1w0q0d0p6h0	9	2219	0.018	0.0003	4.500
MD08 J161133+54160.6	242.88699	54.27652	0.311	c0s0w0q0d0p31h0	43	7532	0.015	0.0006	2.900
MD08 J161341+55420.2	243.42193	55.70259	0.333	c0s0w0q0d0p6h0	6	1070	0.016	0.0002	3.125
MD08 J161128+55510.4	242.86791	55.85700	0.236	c0s0w0q0d0p8h0	11	2618	0.014	0.0003	2.500
MD08 J160946+53460.3	242.44016	53.77208	0.335	c0s0w0q0d0p5h0	5	188	0.029	0.0001	2.000
MD08 J161811+54590.0	244.54397	54.98413	0.429	c0s0w0q0d0p5h0	5	3454	0.018	0.0002	21.000
MD08 J161702+56010.4	244.25836	56.02377	0.410	c0s0w0q0d0p3h0	6	2428	0.022	0.0002	1.875

Continued overleaf...

Table C.3 – continued

Name	RA	dec	cluster_z	cz_type	N_{gal}	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MD08 J160748+55390.7	241.95061	55.66103	0.258	c0s0w0q0d0p4h0	5	0	0.017	0.0001	5.500
MD08 J160240+55110.6	240.66510	55.19346	0.219	c0s1w0q0d0p15h0	19	10320	0.047	0.0008	2.273
MD08 J160539+54560.3	241.41334	54.93803	0.225	c0s0w0q0d0p5h0	6	2453	0.042	0.0002	1.083
MD08 J160439+54310.9	241.16240	54.53131	0.228	c0s1w0q0d0p71h0	82	7226	0.043	0.0010	3.125
MD08 J160730+55020.2	241.87355	55.03657	0.516	c0s0w0q0d0p5h0	6	1227	0.026	0.0001	2.000
MD08 J161415+55470.7	243.56057	55.79446	0.438	c0s0w0q0d0p4h0	6	4546	0.007	0.0002	2.000
MD08 J160502+55410.3	241.25791	55.68907	0.542	c0s0w0q0d0p1h0	5	1335	0.037	0.0001	2.000
MD08 J161956+55010.0	244.98487	55.03321	0.324	c0s0w0q0d0p4h0	5	3930	0.004	0.0001	4.750
MD08 J161506+54250.6	243.77447	54.42654	0.303	c0s0w0q0d0p3h0	5	6443	0.007	0.0002	1.643
MD08 J161319+56190.6	243.32999	56.32709	0.176	c0s0w0q0d0p6h0	6	4751	0.024	0.0002	2.200
MD08 J162121+55160.3	245.33838	55.27129	0.199	c0s0w0q0d0p1h0	5	5540	0.037	0.0002	16.000
MD08 J160131+55500.4	240.37992	55.83941	-	-	10	0	0.036	0.0010	2.000
MD08 J160711+55220.2	241.79618	55.36950	0.626	c0s0w0q0d0p4h0	5	639	0.014	0.0001	3.667
MD08 J161036+54570.4	242.64871	54.95709	0.524	c0s0w0q0d0p6h0	6	519	0.015	0.0001	1.400
MD08 J161420+55480.5	243.58399	55.80778	0.252	c0s0w0q0d0p6h0	6	8131	0.028	0.0001	2.000
MD08 J161958+54570.7	244.99147	54.96185	0.339	c0s0w0q0d0p3h0	5	1389	0.025	0.0001	2.833
MD08 J161117+54170.6	242.81992	54.29279	0.316	c0s0w0q0d0p4h0	6	2377	0.016	0.0002	3.750
MD08 J161249+55020.6	243.20498	55.04409	0.456	c0s0w0q0d0p5h0	7	5651	0.015	0.0002	19.000

Continued overleaf...

Table C.3 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
MD08 J161835+55170.4	244.64582	55.28943	0.241	c0s1w0q0d0p25h0	29	-	0.017	0.0005	3.714
MD08 J161550+54040.9	243.95821	54.08145	0.207	c0s0w0q0d0p4h0	6	2953	0.014	0.0002	3.000
MD08 J161228+54290.1	243.11869	54.48549	0.283	c0s0w0q0d0p4h0	5	15364	0.011	0.0001	2.429
MD08 J160522+55130.1	241.34049	55.21874	0.232	c0s0w0q0d0p15h0	16	2139	0.041	0.0006	3.000
MD08 J161408+53310.9	243.53527	53.53095	0.176	c0s0w0q0d0p5h0	6	1627	0.015	0.0002	2.571

Table C.3: ORCA cluster detections from the 8th Pan-STARRS Medium Deep Survey (MD-8)

C.4 Pan-STARRS Small Area Survey (SAS) Data

Table C.4 – Pan-STARRS-1 SAS catalogue

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
SAS J090638+37180.6	136.65995	37.31057	0.206	c0s2w0q0d0b0p33h0	34	3324	0.044	0.0009	2.231
SAS J092404+37050.2	141.01648	37.08589	0.191	c0s1w0q0d0b0p5h0	6	2705	0.038	0.0003	2.625
SAS J090117+35130.0	135.31932	35.21675	0.191	c0s1w0q0d0b0p6h0	6	2292	0.022	0.0002	1.300
SAS J092029+36410.7	140.12138	36.69507	0.205	c0s1w0q0d0b0p7h0	7	1515	0.044	0.0003	1.714
SAS J092510+37300.1	141.29360	37.50134	0.174	c0s1w0q0d0b0p14h0	15	3261	0.039	0.0005	1.909
SAS J091241+34090.7	138.16903	34.16146	0.198	c0s1w0q0d0b0p8h0	9	2510	0.031	0.0006	3.875
SAS J091608+36330.7	139.03423	36.56142	0.263	c0s0w0q0d0b0p4h0	6	624	0.032	0.0003	1.429
SAS J092040+35310.8	140.16668	35.52972	0.224	c0s1w0q0d0b0p12h0	15	2098	0.040	0.0005	2.727
SAS J092413+37100.3	141.05504	37.17155	0.244	c0s2w0q0d0b0p6h0	6	1465	0.026	0.0002	2.167
SAS J091614+34350.4	139.05758	34.59000	0.237	c0s0w0q0d0b0p8h0	8	1922	0.038	0.0004	2.625
SAS J092232+38340.8	140.63486	38.58060	0.266	c0s0w0q0d0b0p6h0	6	1506	0.038	0.0004	1.556
SAS J091741+36360.5	139.41929	36.60772	0.161	c0s1w0q0d0b0p11h0	12	1856	0.039	0.0006	1.923
SAS J091302+36570.2	138.26030	36.95375	0.202	c0s1w0q0d0b0p6h0	6	765	0.049	0.0002	3.750
SAS J092839+37470.9	142.16385	37.79801	0.231	c0s3w0q0d0b0p48h0	51	3441	0.039	0.0012	4.143
SAS J091456+35500.3	138.73241	35.83800	0.193	c0s0w0q0d0b0p5h0	5	2467	0.034	0.0002	2.500
SAS J092251+35140.5	140.71344	35.24105	0.181	c0s0w0q0d0b0p6h0	6	3617	0.039	0.0004	1.040

Continued overleaf...

Table C.4 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
SAS J092019+37070.4	140.07988	37.12409	0.218	c0s1w0q0d0b0p13h0	14	2652	0.024	0.0007	2.067
SAS J092027+35080.1	140.11272	35.13446	0.210	c0s0w0q0d0b0p7h0	12	2012	0.034	0.0005	2.667
SAS J091452+36340.7	138.71664	36.57791	0.199	c0s1w0q0d0b0p13h0	13	1418	0.037	0.0007	2.417
SAS J091648+37350.9	139.20206	37.59903	0.183	c0s2w0q0d0b0p5h0	7	2787	0.038	0.0004	3.100
SAS J090751+34470.5	136.96136	34.79121	0.258	c0s0w0q0d0b0p5h0	6	2155	0.028	0.0002	1.625
SAS J090928+32580.5	137.36836	32.97442	0.259	c0s1w0q0d0b0p5h0	6	1162	0.041	0.0001	1.750
SAS J091653+34080.0	139.22238	34.13376	0.300	c0s0w0q0d0b0p5h0	6	1891	0.040	0.0002	1.800
SAS J092522+35180.8	141.34358	35.31310	0.258	c0s0w0q0d0b0p5h0	5	932	0.037	0.0002	1.875
SAS J092111+34510.8	140.29466	34.86287	0.228	c0s1w0q0d0b0p8h0	9	3304	0.023	0.0005	7.250
SAS J091612+37220.3	139.05078	37.37130	0.241	c0s1w0q0d0b0p12h0	12	1085	0.048	0.0004	1.909
SAS J091642+36240.5	139.17502	36.40882	0.204	c0s0w0q0d0b0p6h0	6	2580	0.034	0.0002	1.400
SAS J091746+33550.4	139.44363	33.92267	0.213	c0s0w0q0d0b0p5h0	5	3751	0.045	0.0003	5.000
SAS J090445+33170.8	136.18585	33.29749	0.178	c0s2w0q0d0b0p11h0	12	4443	0.039	0.0005	1.263
SAS J091248+38110.2	138.20169	38.18616	0.136	c0s0w0q0d0b0p3h0	6	2810	0.029	0.0003	1.333
SAS J092249+38380.7	140.70478	38.64544	0.277	c0s0w0q0d0b0p8h0	8	1718	0.045	0.0003	1.889
SAS J092311+34180.4	140.79739	34.30646	0.167	c0s0w0q0d0b0p5h0	5	3908	0.050	0.0003	3.143
SAS J091206+33430.5	138.02686	33.72474	0.227	c0s0w0q0d0b0p4h0	5	2561	0.028	0.0002	2.600
SAS J092429+35520.6	141.11959	35.87623	0.204	c0s0w0q0d0b0p4h0	5	1747	0.023	0.0002	3.250

Continued overleaf...

Table C.4 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
SAS J092604+37480.5	141.51470	37.80816	0.342	c0s0w0q0d0b0p5h0	5	732	0.036	0.0003	18.000
SAS J092121+37020.3	140.33931	37.03853	0.243	c0s0w0q0d0b0p10h0	10	2195	0.050	0.0006	1.550
SAS J091020+35260.7	137.58176	35.44421	0.180	c0s1w0q0d0b0p16h0	17	4976	0.035	0.0008	3.500
SAS J090759+34310.7	136.99383	34.52900	0.198	c0s1w0q0d0b0p8h0	10	1649	0.031	0.0004	2.875
SAS J092155+36520.8	140.47886	36.87952	0.202	c0s1w0q0d0b0p4h0	5	2338	0.039	0.0006	7.750
SAS J091544+35500.2	138.93391	35.83740	0.222	c0s0w0q0d0b0p3h0	5	2242	0.044	0.0002	2.333
SAS J092246+36310.2	140.69102	36.52009	0.227	c0s1w0q0d0b0p9h0	9	3069	0.032	0.0005	4.400
SAS J091409+35200.4	138.53834	35.34002	0.219	c0s1w0q0d0b0p8h0	10	1936	0.029	0.0004	3.500
SAS J092245+35020.0	140.68590	35.04977	0.218	c0s1w0q0d0b0p7h0	7	2254	0.044	0.0003	2.000
SAS J091428+37440.5	138.61857	37.74221	0.194	c0s1w0q0d0b0p14h0	14	1485	0.044	0.0005	1.571
SAS J091619+36190.0	139.07809	36.31713	0.198	c0s0w0q0d0b0p5h0	5	2321	0.028	0.0004	1.750
SAS J091655+36400.3	139.22832	36.67105	0.193	c0s0w0q0d0b0p15h0	15	5190	0.039	0.0008	2.231
SAS J092831+35000.4	142.13060	35.00721	0.124	c0s2w0q0d0b0p6h0	6	112	0.024	0.0002	1.375
SAS J091745+36260.7	139.43607	36.44423	0.166	c0s2w0q0d0b0p8h0	8	1735	0.033	0.0003	2.286
SAS J090524+33050.6	136.35167	33.09410	0.178	c0s1w0q0d0b0p4h0	5	2329	0.047	0.0006	2.889
SAS J091642+36360.7	139.17451	36.61232	0.173	c0s0w0q0d0b0p6h0	7	5210	0.044	0.0005	2.700
SAS J091807+36310.9	139.52819	36.53205	0.175	c0s0w0q0d0b0p5h0	6	2067	0.027	0.0005	2.625
SAS J091146+39000.1	137.94279	39.00141	0.153	c0s1w0q0d0b0p6h0	6	1548	0.021	0.0003	3.500

Continued overleaf...

Table C.4 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
SAS J091943+37060.5	139.92956	37.10847	0.117	c0s1w0q0d0b0p9h0	10	3554	0.037	0.0006	2.889
SAS J090844+36330.7	137.18228	36.56099	0.122	c0s0w0q0d0b0p5h0	6	4249	0.024	0.0002	2.500
SAS J090611+35300.0	136.54481	35.51625	0.114	c0s0w0q0d0b0p4h0	5	16803	0.026	0.0004	6.000
SAS J092515+37030.8	141.31322	37.06411	0.216	c0s0w0q0d0b0p6h0	7	1158	0.037	0.0003	2.600
SAS J091407+36530.3	138.52875	36.88914	0.224	c0s0w0q0d0b0p3h0	5	1372	0.007	0.0004	3.714
SAS J091530+38290.4	138.87508	38.48972	0.341	c0s0w0q0d0b0p4h0	5	402	0.030	0.0005	1.824
SAS J090455+35390.5	136.22886	35.65785	0.261	c0s1w0q0d0b0p7h0	7	1545	0.044	0.0004	1.143
SAS J090254+34570.2	135.72638	34.95387	0.304	c0s1w0q0d0b0p6h0	6	849	0.035	0.0002	5.000
SAS J091531+35340.2	138.87714	35.57059	0.268	c0s0w0q0d0b0p4h0	5	1340	0.041	0.0002	7.500
SAS J093234+37540.0	143.14351	37.90005	0.190	c0s1w0q0d0b0p11h0	14	5002	0.035	0.0004	2.429
SAS J091736+38480.9	139.40064	38.81528	0.153	c0s0w0q0d0b0p9h0	9	4819	0.032	0.0006	5.600
SAS J090258+33340.2	135.74228	33.56986	0.264	c0s1w0q0d0b0p5h0	6	3213	0.023	0.0003	2.000
SAS J090504+37530.7	136.26474	37.89517	0.193	c0s0w0q0d0b0p4h0	5	1393	0.031	0.0002	1.700
SAS J090509+34520.6	136.28548	34.87724	0.110	c0s4w0q0d0b0p5h0	5	3212	0.016	0.0002	1.400
SAS J091240+35370.4	138.16526	35.62283	0.130	c0s1w0q0d0b0p4h0	7	1383	0.035	0.0003	1.583
SAS J092609+39360.9	141.53847	39.61426	0.310	c0s0w0q0d0b0p6h0	7	1259	0.029	0.0005	2.222
SAS J091440+38340.4	138.66649	38.57300	0.293	c0s0w0q0d0b0p10h0	13	754	0.034	0.0008	2.188
SAS J090941+37590.4	137.41902	37.98918	0.249	c0s1w0q0d0b0p7h0	8	1924	0.022	0.0003	2.667

Continued overleaf...

Table C.4 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
SAS J091019+37470.5	137.58014	37.79170	0.213	c0s1w0q0d0b0p11h0	11	2651	0.040	0.0004	2.429
SAS J091527+36360.1	138.86253	36.60129	0.201	c0s0w0q0d0b0p6h0	8	4495	0.030	0.0004	6.000
SAS J091712+37090.1	139.30060	37.15146	0.106	c0s1w0q0d0b0p5h0	5	3039	0.030	0.0003	1.889
SAS J093004+38280.5	142.51818	38.47474	0.178	c0s1w0q0d0b0p12h0	14	2969	0.034	0.0010	4.143
SAS J091445+37370.6	138.68917	37.62668	0.194	c0s1w0q0d0b0p5h0	5	0	0.038	0.0002	1.300
SAS J092329+33530.7	140.87156	33.89472	0.233	c0s2w0q0d0b0p14h0	14	2503	0.034	0.0006	2.400
SAS J092258+37320.6	140.74271	37.54315	0.237	c0s0w0q0d0b0p5h0	5	1673	0.031	0.0002	2.833
SAS J092353+35270.7	140.96933	35.46115	0.277	c0s0w0q0d0b0p4h0	5	1340	0.029	0.0003	1.429
SAS J092001+35080.9	140.00471	35.14756	0.209	c0s0w0q0d0b0p5h0	5	2472	0.050	0.0003	3.750
SAS J090811+33450.1	137.04385	33.75226	0.219	c0s0w0q0d0b0p4h0	5	2753	0.021	0.0003	4.000
SAS J091757+36510.8	139.48883	36.86337	0.206	c0s0w0q0d0b0p4h0	5	1447	0.027	0.0004	2.667
SAS J091344+36390.7	138.43522	36.66195	0.386	c0s0w0q0d0b0p4h0	5	1444	0.043	0.0003	1.750
SAS J090533+35060.0	136.38871	35.10026	0.307	c0s1w0q0d0b0p5h0	7	1139	0.044	0.0002	2.000
SAS J092259+33530.6	140.74705	33.89398	0.204	c0s0w0q0d0b0p3h0	5	320	0.031	0.0002	6.500
SAS J092337+35560.7	140.90568	35.94453	0.204	c0s0w0q0d0b0p4h0	5	2798	0.037	0.0003	1.818
SAS J091438+36160.9	138.65656	36.28142	0.112	c0s5w0q0d0b0p9h0	11	1675	0.032	0.0006	3.286
SAS J092727+36080.3	141.86191	36.13826	0.124	c0s0w0q0d0b0p3h0	5	-769	0.020	0.0004	4.750
SAS J091517+36530.6	138.82234	36.89318	0.157	c0s0w0q0d0b0p4h0	5	0	0.032	0.0003	3.000

Continued overleaf...

Table C.4 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
SAS J092203+40100.9	140.51104	40.18166	0.141	c0s1w0q0d0b0p5h0	5	419	0.027	0.0001	3.000
SAS J090902+34300.0	137.26039	34.51600	0.223	c0s1w0q0d0b0p9h0	9	2514	0.036	0.0003	1.556
SAS J091856+38430.4	139.73454	38.72332	0.221	c0s0w0q0d0b0p4h0	5	1477	0.028	0.0002	1.857
SAS J090627+35060.0	136.61138	35.10056	0.255	c0s2w0q0d0b0p5h0	5	1227	0.035	0.0002	3.000
SAS J091953+35510.7	139.96880	35.86183	0.200	c0s1w0q0d0b0p29h0	31	6382	0.039	0.0009	1.867
SAS J091709+35540.1	139.28775	35.90159	0.307	c0s0w0q0d0b0p4h0	8	595	0.021	0.0005	3.167
SAS J090907+36320.3	137.27846	36.53837	0.407	c0s2w0q0d0b0p8h0	9	621	0.038	0.0004	2.000
SAS J090602+35480.2	136.51012	35.80409	0.329	c0s1w0q0d0b0p4h0	5	560	0.051	0.0004	5.000
SAS J091204+33310.4	138.01637	33.52320	0.179	c0s0w0q0d0b0p5h0	5	2272	0.041	0.0003	2.111
SAS J092422+32480.1	141.08968	32.80149	0.307	c0s0w0q0d0b0p5h0	5	1548	0.040	0.0006	1.500
SAS J091051+33560.3	137.71109	33.93787	0.183	c0s3w0q0d0b0p16h0	17	3126	0.046	0.0009	2.778
SAS J091631+39100.9	139.13115	39.18137	0.070	c0s2w0q0d0b0p4h0	5	2986	0.035	0.0002	2.200
SAS J092010+38310.6	140.04369	38.52625	0.377	c0s0w0q0d0b0p8h0	8	666	0.040	0.0007	2.000
SAS J092810+37550.3	142.04099	37.92146	0.377	c0s0w0q0d0b0p5h0	6	2103	0.024	0.0003	2.833
SAS J091135+33270.7	137.89426	33.46191	0.070	c0s0w0q0d0b0p2h0	5	42150	0.017	0.0003	5.500
SAS J091717+35590.5	139.32063	35.99160	0.201	c0s0w0q0d0b0p5h0	5	4757	0.031	0.0001	1.750
SAS J092649+36430.7	141.70488	36.72800	0.242	c0s1w0q0d0b0p6h0	6	2587	0.029	0.0003	1.714
SAS J093040+38330.8	142.66470	38.56375	0.130	c0s1w0q0d0b0p5h0	5	1519	0.021	0.0002	2.167

Continued overleaf...

Table C.4 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
SAS J085926+35160.2	134.85741	35.27005	0.159	c0s1w0q0d0b0p9h0	9	2684	0.039	0.0004	2.364
SAS J090747+38210.0	136.94470	38.35081	0.444	c0s1w0q0d0b0p4h0	6	421	0.034	0.0006	1.727
SAS J090904+35300.6	137.26753	35.51064	0.231	c0s0w0q0d0b0p7h0	8	2214	0.053	0.0008	1.500
SAS J093359+38510.1	143.49600	38.85119	0.230	c0s1w0q0d0b0p17h0	17	2452	0.040	0.0009	3.000
SAS J091546+36020.7	138.94087	36.04558	0.277	c0s0w0q0d0b0p5h0	5	980	0.026	0.0002	2.667
SAS J092822+36460.5	142.08959	36.77543	0.367	c0s1w0q0d0b0p5h0	5	1363	0.026	0.0001	2.000
SAS J092445+39150.7	141.18958	39.26146	0.241	c0s0w0q0d0b0p14h0	15	2166	0.046	0.0005	1.917
SAS J092915+37580.2	142.31095	37.96930	0.224	c0s0w0q0d0b0p6h0	6	3032	0.043	0.0005	2.667
SAS J093544+37520.1	143.93210	37.86812	0.153	c0s1w0q0d0b0p8h0	8	1945	0.029	0.0002	2.500
SAS J091404+37380.1	138.51698	37.63572	0.161	c0s1w0q0d0b0p5h0	6	438	0.035	0.0004	3.600
SAS J090423+32570.9	136.09432	32.96544	0.208	c0s0w0q0d0b0p6h0	6	865	0.041	0.0002	1.444
SAS J092916+37460.8	142.31702	37.78008	0.217	c0s0w0q0d0b0p3h0	5	2782	0.017	0.0004	1.727
SAS J092331+34140.9	140.87879	34.24829	0.214	c0s0w0q0d0b0p5h0	6	1262	0.018	0.0006	5.200
SAS J092344+37240.4	140.93478	37.40690	0.220	c0s0w0q0d0b0p5h0	5	1844	0.042	0.0004	1.333
SAS J085958+35110.4	134.99218	35.19078	0.199	c0s1w0q0d0b0p5h0	6	2487	0.030	0.0004	2.333
SAS J092829+36460.8	142.12219	36.78030	0.141	c0s1w0q0d0b0p5h0	5	5665	0.038	0.0001	3.333
SAS J092539+36270.6	141.41255	36.45959	0.118	c0s2w0q0d0b0p13h0	19	2266	0.041	0.0008	3.000
SAS J092301+37470.4	140.75473	37.78978	0.262	c0s1w0q0d0b0p5h0	5	596	0.015	0.0003	3.250

Continued overleaf...

Table C.4 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
SAS J090809+36490.3	137.03595	36.82143	0.407	c0s0w0q0d0b0p4h0	5	908	0.020	0.0002	2.250
SAS J091744+38500.6	139.43520	38.84275	0.142	c0s1w0q0d0b0p3h0	5	4601	0.036	0.0003	2.286
SAS J091112+38130.3	137.80030	38.22216	0.299	c0s1w0q0d0b0p9h0	10	1366	0.036	0.0006	3.000
SAS J092315+38370.7	140.81073	38.62896	0.148	c0s0w0q0d0b0p6h0	6	1974	0.033	0.0003	2.000
SAS J090949+35280.6	137.45536	35.47661	0.137	c0s2w0q0d0b0p5h0	7	4681	0.025	0.0005	2.600
SAS J092056+33380.6	140.23370	33.64329	0.311	c0s0w0q0d0b0p5h0	5	305	0.013	0.0007	2.250
SAS J091328+37270.5	138.36756	37.45797	0.157	c0s0w0q0d0b0p5h0	5	3426	0.057	0.0004	10.500
SAS J090847+34380.1	137.19458	34.63433	0.352	c0s1w0q0d0b0p11h0	13	700	0.029	0.0006	4.400
SAS J092812+37360.5	142.04952	37.60804	0.146	c0s0w0q0d0b0p5h0	6	2242	0.038	0.0003	2.429
SAS J090950+35320.9	137.45805	35.54793	0.311	c0s1w0q0d0b0p15h0	15	2511	0.038	0.0006	2.300
SAS J091630+36330.7	139.12477	36.56143	0.303	c0s0w0q0d0b0p6h0	6	3376	0.031	0.0007	1.714
SAS J091854+34550.9	139.72694	34.93191	0.220	c0s1w0q0d0b0p10h0	10	2960	0.043	0.0004	2.429
SAS J091646+36390.1	139.19299	36.65202	0.309	c0s0w0q0d0b0p4h0	5	1624	0.035	0.0003	2.250
SAS J092952+37410.2	142.46484	37.68642	0.401	c0s0w0q0d0b0p5h0	6	344	0.050	0.0003	1.267
SAS J091930+38570.8	139.87602	38.96316	0.294	c0s1w0q0d0b0p7h0	8	565	0.035	0.0009	5.286
SAS J090526+33130.4	136.35773	33.22365	0.346	c0s1w0q0d0b0p5h0	5	1424	0.049	0.0002	1.667
SAS J092129+35430.3	140.37141	35.72115	0.367	c0s0w0q0d0b0p3h0	5	797	0.030	0.0002	2.600
SAS J090929+37340.3	137.36929	37.57158	0.238	c0s0w0q0d0b0p5h0	6	2793	0.028	0.0005	3.667

Continued overleaf...

Table C.4 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	bgc	scatter	θ_{80}	θ_{80}/θ_{20}
SAS J090346+34130.9	135.94115	34.23186	0.176	c0s0w0q0d0b0p6h0	8	4083	0.023	0.0004	2.667
SAS J091148+33230.9	137.94820	33.39865	0.249	c0s1w0q0d0b0p6h0	7	1774	0.050	0.0004	2.071
SAS J091809+38510.8	139.53593	38.86291	0.143	c0s1w0q0d0b0p6h0	6	4084	0.027	0.0003	1.300
SAS J090510+34220.1	136.29225	34.36837	0.325	c0s0w0q0d0b0p6h0	6	1442	0.035	0.0002	1.667
SAS J090836+32590.6	137.14900	32.99354	0.311	c0s1w0q0d0b0p8h0	9	1798	0.044	0.0008	2.700
SAS J093131+38200.4	142.88044	38.34041	0.145	c0s1w0q0d0b0p7h0	7	801	0.038	0.0004	2.714
SAS J091014+32580.5	137.55857	32.97565	0.269	c0s0w0q0d0b0p3h0	5	495	0.023	0.0003	1.400
SAS J091153+37240.4	137.97050	37.40683	0.158	c0s0w0q0d0b0p4h0	8	1351	0.031	0.0004	1.700
SAS J090736+38190.0	136.89828	38.33279	0.321	c0s1w0q0d0b0p4h0	5	1041	0.039	0.0003	4.200
SAS J090626+35280.4	136.60852	35.47368	0.084	c0s0w0q0d0b0p6h0	6	43978	0.035	0.0003	2.571
SAS J091056+35310.1	137.73178	35.51856	0.208	c0s0w0q0d0b0p5h0	7	3644	0.023	0.0009	1.692

Table C.4: `ORCA` cluster detections from the Pan-STARRS Small Area Survey (SAS)

C.5 Pan-STARRS MD-5 Data

Table C.5 – Pan-STARRS-1 MD-5 catalogue

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	$A_{gc[1,i]}$	scatter	θ_{80}	θ_{80}/θ_{20}	$\mathrm{M}^{\rm cl}_{\star}$
MD05 J105319+56460.9	163.32709	56.78102	0.250	c0s0w0q0d0b0p3h0	5	0.002	0.028	0.0003	1.786	11.08
MD05 J105731+58000.1	164.38033	58.00157	0.546	c0s0w0q0d0b0p4h0	12	0.006	0.023	0.0002	2.300	12.27
MD05 J103922+57110.7	159.84137	57.19440	0.392	c0s0w0q0d0b0p3h0	5	0.003	0.053	0.0002	2.429	11.30
MD05 J104947+57000.8	162.44542	57.01340	0.160	c0s0w0q0d0b0p3h0	5	0.001	0.045	0.0003	4.000	10.94
MD05 J104017+59110.7	160.07049	59.19534	0.265	c0s0w0q0d0b0p3h0	6	0.000	0.024	0.0003	2.667	11.28
MD05 J104437+58080.8	161.15501	58.14742	0.478	c0s0w0q0d0b0p6h0	11	0.012	0.040	0.0004	2.500	12.02
MD05 J105119+58390.3	162.83029	58.65559	0.465	c0s0w0q0d0b0p7h0	14	0.007	0.017	0.0008	2.188	12.04
MD05 J103920+57140.1	159.83397	57.23573	0.509	c0s0w0q0d0b0p1h0	6	0.006	0.011	0.0002	1.333	11.76
MD05 J104311+58200.4	160.79551	58.34020	0.481	c0s0w0q0d0b0p5h0	10	0.012	0.041	0.0003	2.875	11.87
MD05 J104905+59170.4	162.26983	59.29003	0.538	c0s0w0q0d0b0p3h0	5	0.007	0.009	0.0001	2.667	11.70
MD05 J104605+58310.6	161.52078	58.52696	0.484	c0s1w0q0d0b0p5h0	5	0.001	0.041	0.0002	2.571	11.46
MD05 J103808+58040.3	159.53178	58.07113	0.491	c0s0w0q0d0b0p1h0	6	0.009	0.010	0.0001	2.500	11.68
MD05 J104352+58210.9	160.96863	58.36564	0.534	c0s0w0q0d0b0p3h0	6	0.011	0.029	0.0001	3.333	11.85
MD05 J105346+59140.3	163.44192	59.23830	0.345	c0s0w0q0d0b0p2h0	8	0.009	0.048	0.0009	3.889	11.56
MD05 J104816+57280.2	162.06682	57.46940	0.321	c0s1w0q0d0b0p4h0	10	-0.001	0.039	0.0009	2.231	11.83
MD05 J105215+59080.3	163.06078	59.13765	0.356	c0s1w0q0d0b0p2h0	7	0.009	0.040	0.0005	1.750	11.68

Continued overleaf...

Table C.5 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	$A_{gc[1,i]}$	scatter	θ_{80}	θ_{80}/θ_{20}	$\mathrm{M}^{\rm cl}_{\star}$
MD05 J104810+56440.0	162.04015	56.74917	0.252	c0s0w0q0d0b0p3h0	6	0.002	0.024	0.0002	1.167	11.39
MD05 J105312+56420.6	163.30198	56.70926	0.300	c0s0w0q0d0b0p2h0	7	0.001	0.039	0.0001	3.200	11.34
MD05 J105054+57060.9	162.72329	57.11530	0.400	c0s0w0q0d0b0p4h0	7	0.006	0.027	0.0002	2.714	11.55
MD05 J104240+57430.6	160.66489	57.72586	0.313	c0s0w0q0d0b0p2h0	6	0.012	0.046	0.0002	1.500	11.24
MD05 J104653+58140.3	161.72066	58.23773	0.370	c0s0w0q0d0b0p6h0	9	0.016	0.039	0.0003	3.333	11.60
MD05 J104222+57430.7	160.59176	57.72874	0.416	c0s0w0q0d0b0p4h0	7	0.013	0.030	0.0002	1.375	11.50
MD05 J105557+58210.0	163.98571	58.36637	0.171	c0s0w0q0d0b0p2h0	5	0.000	0.022	0.0000	2.333	11.28
MD05 J104127+57170.0	160.36227	57.28341	0.101	c0s0w0q0d0b0p3h0	6	-0.001	0.037	0.0001	3.750	11.40
MD05 J104731+58490.5	161.87825	58.82531	0.304	c0s1w0q0d0b0p3h0	6	0.002	0.028	0.0001	2.100	11.39
MD05 J104757+57200.6	161.98894	57.34282	0.234	c0s0w0q0d0b0p4h0	5	0.003	0.045	0.0001	4.400	11.13
MD05 J105112+58100.5	162.80105	58.17538	0.377	c0s0w0q0d0b0p5h0	5	0.002	0.032	0.0001	2.500	11.49
MD05 J103932+57000.2	159.88246	57.00387	0.191	c0s0w0q0d0b0p5h0	5	-0.001	0.019	0.0002	2.462	11.42
MD05 J104427+57510.9	161.11085	57.86514	0.090	c0s1w0q0d0b0p5h0	5	0.008	0.030	0.0003	2.778	11.49
MD05 J104336+57220.7	160.90076	57.37787	0.117	c0s4w0q0d0b0p21h0	22	0.026	0.036	0.0008	1.786	12.06
MD05 J104418+58040.1	161.07432	58.06791	0.163	c0s0w0q0d0b0p3h0	6	0.005	0.053	0.0002	1.143	10.80
MD05 J104516+58090.2	161.31525	58.15316	0.118	c0s10w0q0d0b0p66h0	72	0.136	0.039	0.0024	4.429	12.30
MD05 J105051+56550.7	162.71371	56.92804	0.119	c0s0w0q0d0b0p3h0	6	0.020	0.044	0.0017	1.636	10.50
MD05 J103845+57350.6	159.68672	57.59260	0.214	c0s0w0q0d0b0p17h0	28	-0.002	0.042	0.0038	2.438	12.31

Continued overleaf. . .

Table C.5 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	$A_{gc[1,i]}$	scatter	θ_{80}	θ_{80}/θ_{20}	$\mathrm{M}^{\rm cl}_{\star}$
MD05 J104453+56330.9	161.22228	56.56457	0.170	c0s0w0q0d0b0p4h0	5	0.002	0.022	0.0003	3.000	11.10
MD05 J104447+58220.6	161.19718	58.37682	0.207	c0s0w0q0d0b0p1h0	5	0.002	0.013	0.0001	1.400	11.11
MD05 J104546+57120.8	161.43973	57.21276	0.229	c0s0w0q0d0b0p27h0	34	-0.002	0.030	0.0019	1.826	12.18
MD05 J105056+59170.4	162.73436	59.28981	0.189	c0s0w0q0d0b0p11h0	15	0.037	0.032	0.0007	1.545	11.69
MD05 J104020+58570.9	160.08266	58.96465	0.562	c0s0w0q0d0b0p2h0	5	0.007	0.023	0.0003	22.000	11.84
MD05 J104707+59280.0	161.77732	59.46674	-	-	6	0.000	0.021	0.0002	1.467	0.00
MD05 J104441+58370.0	161.16933	58.61705	0.453	c0s0w0q0d0b0p2h0	5	0.000	0.009	0.0006	1.812	11.25
MD05 J105207+57070.8	163.02980	57.13005	0.490	c0s0w0q0d0b0p7h0	12	-0.001	0.035	0.0007	1.500	11.85
MD05 J104526+57560.8	161.35681	57.94734	0.214	c0s0w0q0d0b0p17h0	20	0.011	0.029	0.0007	3.455	12.10
MD05 J104750+57380.6	161.95634	57.64345	0.293	c0s0w0q0d0b0p5h0	9	0.008	0.030	0.0003	2.000	11.51
MD05 J104358+58140.1	160.99061	58.23480	0.175	c0s0w0q0d0b0p6h0	6	0.001	0.040	0.0001	3.000	11.40
MD05 J104114+56530.8	160.30926	56.89669	0.182	c0s1w0q0d0b0p15h0	16	0.008	0.017	0.0004	2.500	11.95
MD05 J104216+58480.2	160.56777	58.80252	0.430	c0s0w0q0d0b0p1h0	5	0.000	0.013	0.0005	1.700	11.24
MD05 J105404+58530.9	163.51858	58.89910	0.375	c0s0w0q0d0b0p6h0	10	-0.002	0.035	0.0027	5.500	12.00
MD05 J104950+58380.0	162.45639	58.63359	0.530	c0s0w0q0d0b0p3h0	6	0.005	0.019	0.0003	4.750	11.70
MD05 J105657+58290.9	164.23633	58.49763	0.419	c0s0w0q0d0b0p5h0	13	0.013	0.039	0.0006	3.429	11.78
MD05 J105755+57220.7	164.47861	57.37857	0.213	c0s0w0q0d0b0p4h0	5	0.010	0.041	0.0004	4.333	11.22
MD05 J105844+58090.0	164.68272	58.16612	0.212	c0s1w0q0d0b0p12h0	16	0.002	0.041	0.0017	3.000	11.62

Continued overleaf...

Table C.5 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	$A_{gc[1,i]}$	scatter	θ_{80}	θ_{80}/θ_{20}	$\mathrm{M}^{\rm cl}_{\star}$
MD05 J104739+59060.6	161.91048	59.10939	0.354	c0s1w0q0d0b0p5h0	10	0.016	0.016	0.0003	1.857	11.90
MD05 J105318+57200.6	163.32428	57.34392	0.340	c0s1w0q0d0b0p2h0	8	0.018	0.013	0.0003	3.200	11.56
MD05 J105326+58410.3	163.35953	58.68859	0.504	c0s0w0q0d0b0p7h0	10	0.007	0.036	0.0006	1.786	11.91
MD05 J104710+57490.4	161.79041	57.82415	0.546	c0s0w0q0d0b0p5h0	5	0.003	0.009	0.0003	3.500	11.73
MD05 J105757+57530.5	164.48642	57.89213	0.093	c0s1w0q0d0b0p10h0	10	0.010	0.038	0.0005	3.600	11.63
MD05 J104400+57010.5	160.99847	57.02447	0.125	c0s0w0q0d0b0p3h0	5	0.006	0.026	0.0002	6.333	11.19
MD05 J105903+57480.5	164.76155	57.80876	0.230	c0s0w0q0d0b0p5h0	6	0.006	0.060	0.0002	1.462	11.65
MD05 J104944+58360.7	162.43221	58.61208	0.144	c0s0w0q0d0b0p6h0	6	0.003	0.029	0.0002	2.143	11.24
MD05 J105627+58170.0	164.11436	58.28384	0.311	c0s0w0q0d0b0p3h0	8	0.004	0.023	0.0003	2.083	11.63
MD05 J104437+56560.7	161.15289	56.94427	0.198	c0s0w0q0d0b0p10h0	17	0.007	0.042	0.0005	2.091	11.99
MD05 J105358+59080.5	163.49060	59.14132	0.197	c0s2w0q0d0b0p27h0	28	0.041	0.037	0.0009	3.250	12.12
MD05 J105846+57460.4	164.69128	57.77254	0.082	c0s0w0q0d0b0p5h0	5	0.001	0.032	0.0002	2.400	10.49
MD05 J104527+56410.1	161.36221	56.68575	0.206	c0s0w0q0d0b0p5h0	5	0.002	0.024	0.0001	2.143	11.38
MD05 J104723+57000.2	161.84469	57.00323	0.269	c0s1w0q0d0b0p23h0	30	0.017	0.027	0.0006	4.000	11.98
MD05 J105113+57110.7	162.80310	57.19526	0.310	c0s0w0q0d0b0p4h0	6	0.008	0.045	0.0007	1.235	11.45
MD05 J103957+58540.8	159.98877	58.91359	0.351	c0s1w0q0d0b0p4h0	6	0.004	0.029	0.0006	1.824	11.37
MD05 J105228+57080.6	163.11755	57.14312	0.305	c0s1w0q0d0b0p4h0	5	0.012	0.039	0.0002	4.000	11.30
MD05 J105152+57090.4	162.96662	57.15651	0.303	c0s0w0q0d0b0p4h0	7	0.018	0.047	0.0002	2.571	11.26

Continued overleaf...

Table C.5 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	$A_{gc[1,i]}$	scatter	θ_{80}	θ_{80}/θ_{20}	$\mathrm{M}^{\rm cl}_{\star}$
MD05 J105631+57370.7	164.12760	57.62760	0.311	c0s0w0q0d0b0p2h0	7	0.010	0.023	0.0002	1.448	11.53
MD05 J104910+59070.2	162.29080	59.11929	0.141	c0s0w0q0d0b0p4h0	5	0.006	0.032	0.0004	2.500	10.78
MD05 J104706+59080.0	161.77362	59.13366	0.211	c0s0w0q0d0b0p7h0	7	0.002	0.028	0.0002	2.400	11.20
MD05 J104103+58170.2	160.26224	58.28601	0.087	c0s2w0q0d0b0p15h0	17	0.018	0.034	0.0009	3.000	11.95
MD05 J105520+58180.4	163.83389	58.30700	0.138	c0s1w0q0d0b0p6h0	7	0.007	0.029	0.0003	2.143	11.47
MD05 J104326+59190.1	160.85914	59.31770	0.201	c0s1w0q0d0b0p4h0	6	0.009	0.050	0.0001	1.125	11.47
MD05 J104130+58200.2	160.37549	58.33622	0.269	c0s0w0q0d0b0p2h0	6	0.000	0.016	0.0001	1.636	11.17
MD05 J104654+59260.9	161.72472	59.44875	0.232	c0s0w0q0d0b0p6h0	9	0.004	0.035	0.0008	4.750	11.55
MD05 J105136+58080.9	162.89944	58.14859	0.494	c0s2w0q0d0b0p7h0	19	0.013	0.034	0.0006	2.143	12.18
MD05 J104206+58410.9	160.52623	58.69905	0.295	c0s0w0q0d0b0p4h0	8	0.004	0.124	0.0003	1.636	11.14
MD05 J105745+58400.0	164.43587	58.66742	0.385	c0s0w0q0d0b0p9h0	16	-0.001	0.035	0.0012	2.667	11.98
MD05 J104930+56550.3	162.37378	56.92086	0.430	c0s0w0q0d0b0p5h0	14	0.003	0.043	0.0006	3.167	11.98
MD05 J104104+56530.1	160.26651	56.88485	0.178	c0s1w0q0d0b0p20h0	25	0.003	0.035	0.0014	3.143	12.08
MD05 J104955+57170.8	162.47916	57.29646	0.334	c0s0w0q0d0b0p4h0	6	0.002	0.030	0.0005	1.455	11.41
MD05 J105237+59190.3	163.15573	59.32189	0.179	c0s0w0q0d0b0p6h0	7	0.012	0.026	0.0010	2.222	11.30
MD05 J105808+57230.3	164.53213	57.38858	0.398	c0s0w0q0d0b0p6h0	14	0.005	0.048	0.0005	2.000	11.75
MD05 J104932+56410.1	162.38214	56.68462	0.065	c0s5w0q0d0b0p9h0	9	0.006	0.035	0.0013	4.200	11.71
MD05 J105537+57160.7	163.90289	57.27889	0.443	c0s0w0q0d0b0p5h0	7	0.008	0.023	0.0004	3.800	11.69

Continued overleaf...

Table C.5 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	$A_{gc[1,i]}$	scatter	θ_{80}	θ_{80}/θ_{20}	$\mathrm{M}^{\rm cl}_{\star}$
MD05 J104839+58230.5	162.16271	58.39163	0.335	c0s0w0q0d0b0p3h0	5	-0.002	0.023	0.0008	7.333	10.91
MD05 J105035+56470.1	162.64696	56.78452	0.207	c0s0w0q0d0b0p5h0	5	0.013	0.031	0.0005	3.500	10.79
MD05 J104656+59020.8	161.73346	59.04648	0.169	c0s0w0q0d0b0p4h0	5	0.003	0.036	0.0002	5.000	11.41
MD05 J104702+56490.3	161.75693	56.82153	0.277	c0s0w0q0d0b0p4h0	6	0.011	0.036	0.0004	1.778	11.35
MD05 J104642+59360.7	161.67391	59.61111	0.213	c0s0w0q0d0b0p24h0	43	0.028	0.041	0.0018	3.875	12.11
MD05 J105402+58470.0	163.50669	58.79925	0.427	c0s0w0q0d0b0p2h0	7	0.006	0.023	0.0001	1.571	11.41
MD05 J104518+56500.6	161.32547	56.84305	0.162	c0s0w0q0d0b0p6h0	6	0.007	0.038	0.0003	1.600	11.35
MD05 J104646+57490.2	161.69030	57.82066	0.354	c0s1w0q0d0b0p6h0	15	-0.001	0.062	0.0012	6.286	11.92
MD05 J105039+57270.1	162.66361	57.45123	0.250	c0s0w0q0d0b0p5h0	5	0.005	0.045	0.0007	4.667	11.28
MD05 J104938+57040.7	162.40910	57.07777	0.417	c0s1w0q0d0b0p16h0	45	-0.002	0.035	0.0030	1.562	12.50
MD05 J104850+59190.7	162.20780	59.32911	0.198	c0s3w0q0d0b0p22h0	25	0.060	0.032	0.0007	3.429	12.05
MD05 J104444+57560.9	161.18179	57.94803	0.146	c0s0w0q0d0b0p5h0	5	0.002	0.024	0.0002	3.667	10.98
MD05 J105217+56540.0	163.07109	56.91596	0.200	c0s0w0q0d0b0p6h0	7	0.014	0.045	0.0007	1.625	11.10
MD05 J104142+58340.2	160.42629	58.57054	0.207	c0s0w0q0d0b0p5h0	5	0.008	0.036	0.0004	3.286	10.91
MD05 J104219+58230.3	160.57931	58.38787	0.494	c0s0w0q0d0b0p2h0	12	-0.001	0.072	0.0012	4.625	11.97
MD05 J105245+58050.6	163.18831	58.09257	0.413	c0s0w0q0d0b0p3h0	6	0.003	0.025	0.0001	1.091	11.48
MD05 J104713+58470.3	161.80345	58.78856	0.342	c0s0w0q0d0b0p3h0	5	0.006	0.042	0.0001	7.000	10.95
MD05 J104955+58050.7	162.48043	58.09505	0.393	c0s0w0q0d0b0p3h0	8	0.005	0.029	0.0007	2.077	11.97

Continued overleaf...

Table C.5 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	$A_{gc[1,i]}$	scatter	θ_{80}	θ_{80}/θ_{20}	$\mathrm{M}^{\rm cl}_{\star}$
MD05 J105234+58540.0	163.14219	58.91591	0.489	c0s0w0q0d0b0p6h0	10	0.000	0.037	0.0007	6.667	12.28
MD05 J104819+58330.7	162.07857	58.56118	0.372	c0s0w0q0d0b0p3h0	5	0.003	0.026	0.0005	1.818	11.34
MD05 J105145+56370.8	162.93606	56.63068	0.258	c0s0w0q0d0b0p9h0	14	-0.002	0.033	0.0017	2.182	11.77
MD05 J104111+56470.2	160.29392	56.78740	0.179	c0s0w0q0d0b0p5h0	6	0.001	0.014	0.0001	1.222	11.09
MD05 J105030+56590.3	162.62467	56.98819	0.256	c0s0w0q0d0b0p3h0	5	-0.002	0.023	0.0009	1.429	11.10
MD05 J104741+58060.9	161.92212	58.11500	0.318	c0s0w0q0d0b0p18h0	29	-0.002	0.038	0.0033	2.615	12.36
MD05 J104451+59340.4	161.21049	59.57252	0.201	c0s0w0q0d0b0p4h0	5	0.007	0.027	0.0004	3.500	11.24
MD05 J103859+57520.4	159.74767	57.87355	0.101	c0s1w0q0d0b0p5h0	5	0.004	0.043	0.0003	1.455	11.31
MD05 J104238+57350.7	160.65632	57.59478	0.079	c0s0w0q0d0b0p6h0	6	0.011	0.027	0.0006	2.857	10.67
MD05 J104046+57100.4	160.19089	57.17254	0.174	c0s0w0q0d0b0p4h0	6	0.004	0.039	0.0003	2.125	10.59
MD05 J104043+57330.1	160.17755	57.55157	0.165	c0s0w0q0d0b0p3h0	5	-0.002	0.045	0.0014	1.875	10.43
MD05 J104053+57200.1	160.22092	57.33566	0.296	c0s1w0q0d0b0p3h0	5	0.006	0.035	0.0004	2.333	11.45
MD05 J104340+59270.5	160.91522	59.45781	0.216	c0s0w0q0d0b0p3h0	5	0.015	0.031	0.0002	3.500	11.26
MD05 J104058+58490.2	160.23972	58.81980	0.458	c0s0w0q0d0b0p3h0	5	-0.002	0.026	0.0011	18.000	11.21
MD05 J105138+56520.8	162.90961	56.87965	0.337	c0s0w0q0d0b0p8h0	20	-0.002	0.039	0.0024	2.600	12.00
MD05 J105521+58430.7	163.83648	58.72759	0.415	c0s0w0q0d0b0p2h0	6	0.003	0.026	0.0004	2.571	11.50
MD05 J104909+57500.5	162.28919	57.84224	0.091	c0s2w0q0d0b0p15h0	15	0.023	0.034	0.0007	2.444	12.14
MD05 J105002+56580.6	162.50981	56.97634	0.291	c0s0w0q0d0b0p4h0	6	0.010	0.025	0.0005	1.769	11.23

Continued overleaf...

Table C.5 – continued

Name	RA	dec	cluster_z	cz_type	$N_{\rm gal}$	$A_{gc[1,i]}$	scatter	θ_{80}	θ_{80}/θ_{20}	$\mathrm{M}^{\rm cl}_{\star}$
MD05 J104028+57220.3	160.11848	57.37098	0.315	c0s0w0q0d0b0p5h0	5	0.005	0.042	0.0001	1.857	11.26
MD05 J104439+56550.4	161.16405	56.92366	0.182	c0s1w0q0d0b0p6h0	7	0.009	0.035	0.0003	1.615	11.47
MD05 J105313+56540.1	163.30597	56.90193	0.550	c0s0w0q0d0b0p5h0	12	-0.001	0.036	0.0015	1.667	12.33
MD05 J105647+59000.6	164.19497	59.01032	0.180	c0s1w0q0d0b0p3h0	6	0.007	0.035	0.0002	1.429	11.48
MD05 J104210+57530.5	160.54040	57.89193	0.465	c0s0w0q0d0b0p4h0	7	0.001	0.027	0.0003	2.600	11.82
MD05 J104650+57180.3	161.70720	57.30526	0.137	c0s0w0q0d0b0p4h0	5	0.002	0.042	0.0012	2.875	10.76
MD05 J104752+57200.6	161.96807	57.34396	0.344	c0s0w0q0d0b0p5h0	8	0.003	0.037	0.0003	6.500	11.67
MD05 J104837+59350.4	162.15432	59.59055	0.391	c0s1w0q0d0b0p11h0	21	-0.001	0.040	0.0023	4.333	11.99
MD05 J104140+58010.2	160.41778	58.02081	0.274	c0s0w0q0d0b0p3h0	5	0.009	0.029	0.0004	2.500	11.27

Table C.5: ORCA cluster detections from the 5th Pan-STARRS Medium Deep Survey (MD-5). All quantities except the 7$\rm th$ and final column are as defined in §3.4.1. The optical richness estimator $A_{gc[1,i]}$ is defined in §4.4.2, and $\mathrm{M}^{\rm cl}_{\star}$ (the cluster stellar mass in units $log_{10}h^{-1}M_{\odot}$) is estimated from the Pan-STARRS y-band, as described in §5.5.

C.6 SDSS DR7 slope-normalisation spline data

Colour	Knots	Coefficients	degree
g-r	0.74387715	-0.03873526	3
	0.74387715	-0.04950758	
	0.74387715	-0.04656256	
	0.74387715	-0.04720442	
	1.80056787	0.	
	1.80056787	0.	
	1.80056787	0.	
	1.80056787	0.	
r-i	0.24445163	-0.02717082	3
	0.24445163	-0.00116641	
	0.24445163	-0.02491031	
	0.24445163	-0.04442966	
	1.3142992	0.	
	1.3142992	0.	
	1.3142992	0.	
	1.3142992	0.	
i-z	0.18375215	-0.02593966	3
	0.18375215	-0.0261061	
	0.18375215	0.00612762	
	0.18375215	-0.04328547	
	0.66862139	0.	
	0.66862139	0.	
	0.66862139	0.	
	0.66862139	0.	

Table C.6: Based on Stripe 82 `ORCA` cluster detections, this Table contains the data required to reconstruct the splines used to quantify the $c_{\mathrm{m}20} - \beta$ relation for the *g-r* , *r-i* and *i-z* colours. These splines may be evaluated with the `FITPACK` FORTRAN module (e.g. `splev.f`).

Bibliography

K. N. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. Allende Prieto, D. An, K. S. J. Anderson, and Anderson, S. F et al. The Seventh Data Release of the Sloan Digital Sky Survey. ApJS, 182:543–558, June 2009.

G. O. Abell. The Distribution of Rich Clusters of Galaxies. ApJS, 3:211, May 1958.

G. O. Abell, H. G. Corwin, Jr., and R. P. Olowin. A catalog of rich clusters of galaxies. ApJS, 70:1–138, May 1989.

S. Andreon, H. Quintana, M. Tajer, G. Galaz, and J. Surdej. The Butcher-Oemler effect at z˜ 0.35: a change in perspective. MNRAS, 365:915–928, January 2006.

R. E. Angulo, C. M. Baugh, C. S. Frenk, and C. G. Lacey. The detectability of baryonic acoustic oscillations in future galaxy surveys. MNRAS, 383:755–776, January 2008.

M. A. Aragón-Calvo, B. J. T. Jones, R. van de Weygaert, and J. M. van der Hulst. The multiscale morphology filter: identifying and extracting spatial patterns in the galaxy distribution. A&A, 474:315–338, October 2007.

M. A. Aragón-Calvo, E. Platen, R. van de Weygaert, and A. S. Szalay. The Spine of the Cosmic Web. ApJ, 723:364–382, November 2010.

C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa. The quickhull algorithm for convex hulls. *ACM TRANSACTIONS ON MATHEMATICAL SOFTWARE*, 22(4): 469–483, 1996.

W. A. Barkhouse, P. J. Green, A. Vikhlinin, D.-W. Kim, D. Perley, R. Cameron, J. Silverman, and Mossman A. et al. ChaMP Serendipitous Galaxy Cluster Survey. ApJ, 645: 955–976, July 2006.

C. D. Barr. *Applications of Voronoi tessellations in point pattern analyis*. PhD thesis, University of California, Los Angeles, 2008.

J. D. Barrow, S. P. Bhavsar, and D. H. Sonoda. Minimal spanning trees, filaments and galaxy clustering. MNRAS, 216:17–35, September 1985.

B. Bassett, A. Becker, D. Bizyaev, H. Brewington, C. Choi, and Cinabro, D. et al. Supernovae 2007pn-2007qb. *Central Bureau Electronic Telegrams*, 1135:1–+, November 2007.

C. M. Baugh, D. J. Croton, E. Gaztañaga, P. Norberg, M. Colless, I. K. Baldry, J. Bland-Hawthorn, and Bridges, T. et al. The 2dF Galaxy Redshift Survey: hierarchical galaxy clustering. MNRAS, 351:L44–L48, June 2004.

C. M. Baugh, C. G. Lacey, C. S. Frenk, G. L. Granato, L. Silva, A. Bressan, A. J. Benson, and S. Cole. Can the faint submillimetre galaxies be explained in the Λ cold dark matter model? MNRAS, 356:1191–1200, January 2005.

A. J. Benson and R. Bower. Galaxy formation spanning cosmic history. MNRAS, 405: 1573–1623, July 2010.

A. A. Berlind, J. Frieman, D. H. Weinberg, M. R. Blanton, M. S. Warren, K. Abazajian, R. Scranton, and Hogg D. W. et al. Percolation Galaxy Groups and Clusters in the SDSS Redshift Survey: Identification, Catalogs, and the Multiplicity Function. ApJS, 167:1–25, November 2006.

E. Bertin and S. Arnouts. SExtractor: Software for source extraction. A&AS, 117:393–404, June 1996.

M. S. Bessell. The Hipparcos and Tycho Photometric System Passbands. PASP, 112: 961–965, July 2000.

S. P. Bhavsar and J. D. Barrow. Percolation analyses of observed and simulated galaxy clustering. MNRAS, 205:61P–66P, November 1983.

C. Bildfell, H. Hoekstra, A. Babul, and A. Mahdavi. Resurrecting the red from the dead: optical properties of BCGs in X-ray luminous clusters. MNRAS, 389:1637–1654, October 2008.

M. Blair and G. Gilmore. Color Equations for the United Kingdom Schmidt Telescope. PASP, 94:742, August 1982.

M. R. Blanton and S. Roweis. K-Corrections and Filter Transformations in the Ultraviolet, Optical, and Near-Infrared. AJ, 133:734–754, February 2007.

H. Böhringer, W. Voges, J. P. Huchra, B. McLean, R. Giacconi, P. Rosati, R. Burg, and Mader, J. et al. The Northern ROSAT All-Sky (NORAS) Galaxy Cluster Survey. I. X-Ray Properties of Clusters Detected as Extended X-Ray Sources. ApJS, 129:435–474, August 2000.

M. Bolzonella, J.-M. Miralles, and R. Pelló. Photometric redshifts based on standard SED fitting procedures. A&A, 363:476–492, November 2000.

J. R. Bond, L. Kofman, and D. Pogosyan. How filaments of galaxies are woven into the cosmic web. Nature, 380:603–606, April 1996.

N. A. Bond, M. A. Strauss, and R. Cen. Crawling the cosmic network: exploring the morphology of structure in the galaxy distribution. MNRAS, 406:1609–1628, August 2010.

R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. MNRAS, 370:645–655, August 2006.

R. G. Bower, J. R. Lucey, and R. S. Ellis. Precision Photometry of Early Type Galaxies in the Coma and Virgo Clusters - a Test of the Universality of the Colour / Magnitude Relation - Part Two - Analysis. MNRAS, 254:601, February 1992.

G. B. Brammer, P. G. van Dokkum, and P. Coppi. EAZY: A Fast, Public Photometric Redshift Code. ApJ, 686:1503–1513, October 2008.

M. Brodwin, J. Ruel, P. A. R. Ade, K. A. Aird, K. Andersson, M. L. N. Ashby, M. Bautz, and Bazin G. et al. SPT-CL J0546-5345: A Massive z>1 Galaxy Cluster Selected Via the Sunyaev-Zel'dovich Effect with the South Pole Telescope. ApJ, 721:90–97, September 2010.

G. Bruzual and S. Charlot. Spectral evolution of stellar populations using isochrone synthesis. ApJ, 405:538–553, March 1993.

R. A. Burenin, A. Vikhlinin, A. Hornstrup, H. Ebeling, H. Quintana, and A. Meshcheryakov. The 400 Square Degree ROSAT PSPC Galaxy Cluster Survey: Catalog and Statistical Calibration. ApJS, 172:561–582, October 2007.

H. Butcher and A. Oemler, Jr. The evolution of galaxies in clusters. I - ISIT photometry of C1 0024+1654 and 3C 295. ApJ, 219:18–30, January 1978.

Y.-C. Cai, R. E. Angulo, C. M. Baugh, S. Cole, C. S. Frenk, and A. Jenkins. Mock galaxy redshift catalogues from simulations: implications for Pan-STARRS1. MNRAS, 395: 1185–1203, May 2009.

D. Calzetti, L. Armus, R. C. Bohlin, A. L. Kinney, J. Koornneef, and T. Storchi-Bergmann. The Dust Content and Opacity of Actively Star-forming Galaxies. ApJ, 533:682–695, April 2000.

P. L. Capak, D. Riechers, N. Z. Scoville, C. Carilli, P. Cox, R. Neri, B. Robertson, and Salvato, M. et al. A massive protocluster of galaxies at a redshift of z$\sim$5.3. Nature, 470: 233–235, February 2011.

S. M. Carroll. The Cosmological Constant. *Living Reviews in Relativity*, 4:1–+, February 2001.

M. Casali, A. Adamson, C. Alves de Oliveira, O. Almaini, K. Burch, and Chuter, T. et al. The UKIRT wide-field camera. A&A, 467:777–784, May 2007.

R. Cen and J. P. Ostriker. Where Are the Baryons? II. Feedback Effects. ApJ, 650:560–572, October 2006.

D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky. A Direct Empirical Proof of the Existence of Dark Matter. ApJ, 648:L109–L113, September 2006.

M. Cohen, W. A. Wheaton, and S. T. Megeath. Spectral Irradiance Calibration in the Infrared. XIV. The Absolute Calibration of 2MASS. AJ, 126:1090–1096, August 2003.

J. M. Colberg. Quantifying cosmic superstructures. MNRAS, 375:337–347, February 2007.

J. M. Colberg, K. S. Krughoff, and A. J. Connolly. Intercluster filaments in a ΛCDM Universe. MNRAS, 359:272–282, May 2005.

J. M. Colberg, S. D. M. White, A. Jenkins, and F. R. Pearce. Linking cluster formation to large-scale structure. MNRAS, 308:593–598, September 1999.

S. Cole and C. Lacey. The structure of dark matter haloes in hierarchical clustering models. MNRAS, 281:716, July 1996.

S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk. Hierarchical galaxy formation. MNRAS, 319:168–204, November 2000.

S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, and Bland-Hawthorn, J. et al. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. MNRAS, 362:505–534, September 2005.

G. D. Coleman, C.-C. Wu, and D. W. Weedman. Colors and magnitudes predicted for high redshift galaxies. ApJS, 43:393–416, July 1980.

M. Colless, G. Dalton, S. Maddox, W. Sutherland, P. Norberg, S. Cole, J. Bland-Hawthorn, and Bridges, T. et al. The 2dF Galaxy Redshift Survey: spectra and redshifts. MNRAS, 328:1039–1063, December 2001.

A. A. Collister and O. Lahav. ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. PASP, 116:345–351, April 2004.

C. Conroy, R. H. Wechsler, and A. V. Kravtsov. Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time. ApJ, 647:201–214, August 2006.

K. E. K. Coppin, J. E. Geach, I. Smail, L. Dunne, A. C. Edge, R. J. Ivison, S. Maddox, and Auld, R. et al. Herschel-ATLAS: detection of a far-infrared population around galaxy clusters. *ArXiv e-prints*, May 2011.

R. A. Crain, T. Theuns, C. Dalla Vecchia, V. R. Eke, C. S. Frenk, A. Jenkins, S. T. Kay, and Peacock, J. A. et al. Galaxies-intergalactic medium interaction calculation - I. Galaxy formation as a function of large-scale environment. MNRAS, 399:1773–1794, November 2009.

C. S. Crawford, A. C. Edge, A. C. Fabian, S. W. Allen, H. Bohringer, H. Ebeling, R. G. McMahon, and W. Voges. Optical spectroscopy of the ROSAT X-ray brightest clusters - II. MNRAS, 274:75–84, May 1995.

S. M. Croom, G. T. Richards, T. Shanks, B. J. Boyle, R. G. Sharp, J. Bland-Hawthorn, T. Bridges, and Brunner, R. J. et al. The 2dF-SDSS QSO survey (Croom+, 2009). *VizieR Online Data Catalog*, 739:20019–+, October 2009.

D. J. Croton, M. Colless, E. Gaztañaga, C. M. Baugh, P. Norberg, I. K. Baldry, J. Bland-Hawthorn, and Bridges, T. et al. The 2dF Galaxy Redshift Survey: voids and hierarchical scaling models. MNRAS, 352:828–836, August 2004a.

D. J. Croton, G. R. Farrar, P. Norberg, M. Colless, J. A. Peacock, I. K. Baldry, C. M. Baugh, and Bland-Hawthorn, J. et al. The 2dF Galaxy Redshift Survey: luminosity functions by density environment and galaxy type. MNRAS, 356:1155–1167, January 2005.

D. J. Croton, E. Gaztañaga, C. M. Baugh, P. Norberg, M. Colless, I. K. Baldry, J. Bland-Hawthorn, and Bridges, T. et al. The 2dF Galaxy Redshift Survey: higher-order galaxy correlation functions. MNRAS, 352:1232–1244, August 2004b.

I. Csabai, T. Budavári, A. J. Connolly, A. S. Szalay, Z. Győry, N. Benítez, J. Annis, and Brinkmann, J. et al. The Application of Photometric Redshifts to the SDSS Early Data Release. AJ, 125:580–592, February 2003.

G. B. Dalton, S. J. Maddox, W. J. Sutherland, and G. Efstathiou. The APM Galaxy Survey - V. Catalogues of galaxy clusters. MNRAS, 289:263–284, August 1997.

R. Davé, R. Cen, J. P. Ostriker, G. L. Bryan, L. Hernquist, N. Katz, D. H. Weinberg, and Norman, M. L. et al. Baryons in the Warm-Hot Intergalactic Medium. ApJ, 552:473–483, May 2001.

M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. ApJ, 292:371–394, May 1985.

S. de la Torre, B. Meneux, G. De Lucia, J. Blaizot, O. Le Fèvre, B. Garilli, O. Cucciati, and Mellier, Y. et al. Comparison of the VIMOS-VLT Deep Survey with the Munich semi-analytical model. I. Magnitude counts, redshift distribution, colour bimodality, and galaxy clustering. A&A, 525:A125+, January 2011.

V. de Lapparent, M. J. Geller, and J. P. Huchra. A slice of the universe. ApJ, 302:L1–L5, March 1986.

R. De Propris, W. J. Couch, M. Colless, G. B. Dalton, C. Collins, C. M. Baugh, J. Bland-Hawthorn, and Bridges, T. et al. The 2dF Galaxy Redshift Survey: a targeted study of catalogued clusters of galaxies. MNRAS, 329:87–101, January 2002.

G. de Vaucouleurs. Evidence for a local super,galaxy. AJ, 58:30, February 1953a.

G. de Vaucouleurs. On the distribution of mass and luminosity in elliptical galaxies. MNRAS, 113:134, 1953b.

A. Dekel, F. Stoehr, G. A. Mamon, T. J. Cox, G. S. Novak, and J. R. Primack. Lost and found dark matter in elliptical galaxies. Nature, 437:707–710, September 2005.

Paul Dierckx. *Curve and surface fitting with splines*. Oxford University Press, Inc., New York, NY, USA, 1993. ISBN 0-19-853441-8.

J. P. Dietrich, T. Erben, G. Lamer, P. Schneider, A. Schwope, and Hartlap, J. et al. BLOX: the Bonn lensing, optical, and X-ray selected galaxy clusters. I. Cluster catalog construction. A&A, 470:821–834, August 2007.

J. P. Dietrich, P. Schneider, D. Clowe, E. Romano-Díaz, and J. Kerp. Weak lensing study of dark matter filaments and application to the binary cluster A 222 and A 223. A&A, 440:453–471, September 2005.

M. Doherty, M. Tanaka, C. De Breuck, C. Ly, T. Kodama, J. Kurk, N. Seymour, and Vernet, J. et al. Optical and near-IR spectroscopy of candidate red galaxies in two z$\sim$2.5 proto-clusters. A&A, 509:A83+, January 2010.

M. J. Drinkwater, R. J. Jurek, C. Blake, D. Woods, K. A. Pimbblet, K. Glazebrook, R. Sharp, and Pracy, M. B. et al. The WiggleZ Dark Energy Survey: survey design and first data release. MNRAS, 401:1429–1452, January 2010.

Charles Duyckaerts and Gilles Godefroy. Voronoi tessellation to study the numerical density and the spatial distribution of neurones. *Journal of Chemical Neuroanatomy*, 20(1):83 – 92, 2000. ISSN 0891-0618. URL http://www.sciencedirect.com/science/article/B6T02-43HDYPB-9/2/b607f1fce2f

H. Ebeling, A. C. Edge, S. W. Allen, C. S. Crawford, A. C. Fabian, and J. P. Huchra. The ROSAT Brightest Cluster Sample - IV. The extended sample. MNRAS, 318:333–340, October 2000.

H. Ebeling, A. C. Edge, H. Bohringer, S. W. Allen, C. S. Crawford, A. C. Fabian, W. Voges, and J. P. Huchra. The ROSAT Brightest Cluster Sample - I. The compilation of the sample and the cluster log N-log S distribution. MNRAS, 301:881–914, December 1998.

H. Ebeling and G. Wiedenmann. Detecting structure in two dimensions combining Voronoi tessellation and percolation. Phys. Rev. E, 47:704–710, January 1993.

J. Einasto, G. Hütsi, M. Einasto, E. Saar, D. L. Tucker, V. Müller, P. Heinämäki, and S. S. Allam. Clusters and superclusters in the Sloan Digital Sky Survey. A&A, 405:425–443, July 2003.

D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, and Seo, H.-J. et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. ApJ, 633:560–574, November 2005.

V. R. Eke, C. M. Baugh, S. Cole, C. S. Frenk, P. Norberg, J. A. Peacock, I. K. Baldry, and Bland-Hawthorn, J. et al. Galaxy groups in the 2dFGRS: the group-finding algorithm and the 2PIGG catalogue. MNRAS, 348:866–878, March 2004a.

V. R. Eke, C. S. Frenk, C. M. Baugh, S. Cole, P. Norberg, J. A. Peacock, I. K. Baldry, and Bland-Hawthorn, J. et al. Galaxy groups in the Two-degree Field Galaxy Redshift Survey: the luminous content of the groups. MNRAS, 355:769–784, December 2004b.

P. Erdoğdu, O. Lahav, S. Zaroubi, G. Efstathiou, S. Moody, J. A. Peacock, M. Colless, and Baldry, I. K. et al. The 2dF Galaxy Redshift Survey: Wiener reconstruction of the cosmic web. MNRAS, 352:939–960, August 2004.

A. E. Evrard, T. J. MacFarland, H. M. P. Couchman, J. M. Colberg, N. Yoshida, S. D. M. White, A. Jenkins, and Frenk, C.S et al. Galaxy Clusters in Hubble Volume Simulations: Cosmological Constraints from Sky Survey Populations. ApJ, 573:7–36, July 2002.

J. E. Felten and P. Morrison. Omnidirectional Inverse Compton and Synchrotron Radiation from Cosmic Distributions of Fast Electrons and Thermal Photons. ApJ, 146:686, December 1966.

A. S. Font, R. G. Bower, I. G. McCarthy, A. J. Benson, C. S. Frenk, J. C. Helly, C. G. Lacey, and Baugh, C. M. et al. The colours of satellite galaxies in groups and clusters. MNRAS, 389:1619–1629, October 2008.

J. E. Forero-Romero, Y. Hoffman, S. Gottlöber, A. Klypin, and G. Yepes. A dynamical classification of the cosmic web. MNRAS, 396:1815–1824, July 2009.

J. A. Frieman, M. S. Turner, and D. Huterer. Dark Energy and the Accelerating Universe. ARA&A, 46:385–432, September 2008.

R. R. Gal, R. R. de Carvalho, S. C. Odewahn, S. G. Djorgovski, and V. E. Margoniner. The Northern Sky Optical Cluster Survey. I. Detection of Galaxy Clusters in DPOSS. AJ, 119:12–20, January 2000.

L. Gao, V. Springel, and S. D. M. White. The age dependence of halo clustering. MNRAS, 363:L66–L70, October 2005.

L. Gao and S. D. M. White. Assembly bias in the clustering of dark matter haloes. MNRAS, 377:L5–L9, April 2007.

G. Gavazzi, A. Boselli, A. Donati, P. Franzetti, and M. Scodeggio. Introducing GOLDMine: A new galaxy database on the WEB. A&A, 400:451–455, March 2003.

E. Gaztañaga, P. Norberg, C. M. Baugh, and D. J. Croton. Statistical analysis of galaxy surveys - II. The three-point galaxy correlation function measured from the 2dFGRS. MNRAS, 364:620–634, December 2005.

J. E. Geach, D. N. A. Murphy, and R. G. Bower. 4098 galaxy clusters to z$\sim$0.6 in the Sloan Digital Sky Survey equatorial Stripe 82. MNRAS, 413:3059–3067, June 2011.

James Geach, Mark Lacy, Duncan Farrah, Matt Jarvis, David Murphy, Richard Bower, Jean-Christophe Mauduit, and Gonzlez-Solares, Eduardo et al. The Spitzer Extragalactic Representative Volume Survey: The cluster galaxy stellar mass function at z$\sim$1. MNRAS, (submitted), July 2010.

M. J. Geller and J. P. Huchra. Mapping the universe. *Science*, 246:897–903, November 1989.

B. F. Gerke, J. A. Newman, S. M. Faber, M. C. Cooper, D. J. Croton, M. Davis, C. N. A. Willmer, and Yan, R. et al. The DEEP2 galaxy redshift survey: the evolution of the blue fraction in groups and the field. MNRAS, 376:1425–1444, April 2007.

D. G. Gilbank and M. L. Balogh. Tracking down a critical halo mass for killing galaxies through the growth of the red sequence. MNRAS, 385:L116–L119, March 2008.

M. D. Gladders, O. Lopez-Cruz, H. K. C. Yee, and T. Kodama. The Slope of the Cluster Elliptical Red Sequence: A Probe of Cluster Evolution. ApJ, 501:571, July 1998.

M. D. Gladders and H. K. C. Yee. A New Method For Galaxy Cluster Detection. I. The Algorithm. AJ, 120:2148–2162, October 2000.

M. D. Gladders and H. K. C. Yee. The Red-Sequence Cluster Survey. I. The Survey and Cluster Catalogs for Patches RCS 0926+37 and RCS 1327+29. ApJS, 157:1–29, March 2005.

M. D. Gladders, H. K. C. Yee, S. Majumdar, L. F. Barrientos, H. Hoekstra, P. B. Hall, and L. Infante. Cosmological Constraints from the Red-Sequence Cluster Survey. ApJ, 655: 128–134, January 2007.

R. Gobat, E. Daddi, M. Onodera, A. Finoguenov, A. Renzini, N. Arimoto, R. Bouwens, and Brusa M. et al. A mature cluster with X-ray emission at z = 2.07. A&A, 526:A133+, February 2011.

T. Goto, M. Sekiguchi, R. C. Nichol, N. A. Bahcall, R. S. J. Kim, J. Annis, Ž. Ivezić, and Brinkmann, J. et al. The Cut-and-Enhance Method: Selecting Clusters of Galaxies from the Sloan Digital Sky Survey Commissioning Data. AJ, 123:1807–1825, April 2002.

J. R. I. Gott, M. Jurić, D. Schlegel, F. Hoyle, M. Vogeley, M. Tegmark, N. Bahcall, and J. Brinkmann. A Map of the Universe. ApJ, 624:463–484, May 2005.

J. E. Gunn, M. Carr, C. Rockosi, M. Sekiguchi, K. Berry, B. Elms, E. de Haas, and Ž et al. Ivezić. The Sloan Digital Sky Survey Photometric Camera. AJ, 116:3040–3081, December 1998.

O. Hahn, C. Porciani, C. M. Carollo, and A. Dekel. Properties of dark matter haloes in clusters, filaments, sheets and voids. MNRAS, 375:489–499, February 2007.

N. C. Hambly, R. S. Collins, N. J. G. Cross, R. G. Mann, M. A. Read, and Sutorius, E. T. W. et al. The WFCAM Science Archive. MNRAS, 384:637–662, February 2008.

J. Hao, B. P. Koester, T. A. Mckay, E. S. Rykoff, E. Rozo, A. Evrard, J. Annis, and Becker, M. Precision Measurements of the Cluster Red Sequence Using an Error-Corrected Gaussian Mixture Model. ApJ, 702:745–758, September 2009.

P. C. Hewett, S. J. Warren, S. K. Leggett, and S. T. Hodgkin. The UKIRT Infrared Deep Sky Survey ZY JHK photometric system: passbands and synthetic colours. MNRAS, 367:454–468, April 2006.

S. Hilbert and S. D. M. White. Abundances, masses and weak-lensing mass profiles of galaxy clusters as a function of richness and luminosity in ΛCDM cosmologies. MNRAS, 404:486–501, May 2010.

M. Hilton, E. Lloyd-Davies, S. A. Stanford, J. P. Stott, C. A. Collins, A. K. Romer, M. Hosmer, and Hoyle, B. et al. The XMM Cluster Survey: Active Galactic Nuclei and Starburst Galaxies in XMMXCS J2215.9-1738 at z = 1.46. ApJ, 718:133–147, July 2010.

A. D. Hincks, V. Acquaviva, P. A. R. Ade, P. Aguirre, M. Amiri, J. W. Appel, L. F. Barrientos, and Battistelli E. S. et al. The Atacama Cosmology Telescope (ACT): Beam Profiles and First SZ Cluster Maps. ApJS, 191:423–438, December 2010.

G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, and Gold, B. et al. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data Processing, Sky Maps, and Basic Results. ApJS, 180:225–245, February 2009.

S. T. Hodgkin, M. J. Irwin, P. C. Hewett, and S. J. Warren. The UKIRT wide field camera ZYJHK photometric system: calibration from 2MASS. MNRAS, 394:675–692, April 2009.

J. P. Huchra and M. J. Geller. Groups of galaxies. I - Nearby groups. ApJ, 257:423–437, June 1982.

J. C. Jackson. A critique of Rees's theory of primordial gravitational radiation. MNRAS, 156:1P, 1972.

N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, and Hinshaw, G. et al. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results. ApJS, 192:14, February 2011.

A. Jenkins, C. S. Frenk, F. R. Pearce, P. A. Thomas, J. M. Colberg, S. D. M. White, H. M. P. Couchman, and Peacock, J. A. et al. Evolution of Structure in Cold Dark Matter Universes. ApJ, 499:20, May 1998.

A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida. The mass function of dark matter haloes. MNRAS, 321: 372–384, February 2001.

W. C. Jones, P. A. R. Ade, J. J. Bock, J. R. Bond, J. Borrill, A. Boscaleri, P. Cabella, and Contaldi, C. R. et al. A Measurement of the Angular Power Spectrum of the CMB Temperature Anisotropy from the 2003 Flight of BOOMERANG. ApJ, 647:823–832, August 2006.

E. Jullo, J.-P. Kneib, M. Limousin, Á. Elíasdóttir, P. J. Marshall, and T. Verdugo. A Bayesian approach to strong lensing modelling of galaxy clusters. *New Journal of Physics*, 9:447, December 2007.

N. Kaiser. Clustering in real space and in redshift space. MNRAS, 227:1–21, July 1987.

P. Katgert, A. Mazure, R. den Hartog, C. Adami, A. Biviano, and J. Perea. The ESO Nearby Abell Cluster Survey. V. The catalogue: Contents and instructions for use. A&AS, 129:399–412, April 1998.

T. Kiang. Random Fragmentation in Two and Three Dimensions. *Zeitschrift fur Astrophysik*, 64:433, 1966.

H.-S. Kim, C. M. Baugh, S. Cole, C. S. Frenk, and A. J. Benson. Modelling galaxy clustering: is new physics needed in galaxy formation models? MNRAS, 400:1527–1540, December 2009.

J.-W. Kim, A. C. Edge, D. A. Wake, and J. P. Stott. Clustering properties of high-redshift red galaxies in SA22 from the UKIDSS Deep eXtragalactic Survey. MNRAS, 410:241–256, January 2011.

T. Kodama and N. Arimoto. Origin of the colour-magnitude relation of elliptical galaxies. A&A, 320:41–53, April 1997.

T. Kodama and R. Bower. The K_s-band luminosity and stellar mass functions of galaxies in z˜ 1 clusters. MNRAS, 346:1–12, November 2003.

T. Kodama, I. Tanaka, M. Kajisawa, J. Kurk, B. Venemans, C. De Breuck, J. Vernet, and C. Lidman. The first appearance of the red sequence of galaxies in proto-clusters at 2 $<\sim z <\sim$ 3. MNRAS, 377:1717–1725, June 2007.

B. P. Koester, T. A. McKay, J. Annis, R. H. Wechsler, A. Evrard, L. Bleem, M. Becker, and Johnston, D. et al. A MaxBCG Catalog of 13,823 Galaxy Clusters from the Sloan Digital Sky Survey. ApJ, 660:239–255, May 2007a.

B. P. Koester, T. A. McKay, J. Annis, R. H. Wechsler, A. E. Evrard, E. Rozo, L. Bleem, and Sheldon, E. S. et al. MaxBCG: A Red-Sequence Galaxy Cluster Finder. ApJ, 660: 221–238, May 2007b.

E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, and Larson, D. et al. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation. ApJS, 180:330–376, February 2009.

M. Kriek, P. G. van Dokkum, I. Labbé, M. Franx, G. D. Illingworth, D. Marchesini, and R. F. Quadri. An Ultra-Deep Near-Infrared Spectrum of a Compact Quiescent Galaxy at z = 2.2. ApJ, 700:221–231, July 2009.

T. F. Laganá, R. S. de Souza, and G. R. Keller. On the influence of non-thermal pressure on the mass determination of galaxy clusters. A&A, 510:A76+, February 2010.

O. Lahav and Y. Suto. Measuring our Universe from Galaxy Redshift Surveys. *Living Reviews in Relativity*, 7:8, July 2004.

A. Lawrence, S. J. Warren, O. Almaini, A. C. Edge, N. C. Hambly, and Jameson, R. F. et al. The UKIRT Infrared Deep Sky Survey (UKIDSS). MNRAS, 379:1599–1617, August 2007.

A. D. Lewis, D. A. Buote, and J. T. Stocke. Chandra Observations of A2029: The Dark Matter Profile Down to below $0.01r_{vir}$ in an Unusually Relaxed Cluster. ApJ, 586:135–142, March 2003.

Andrew Liddle. *An Introduction to Modern Cosmology*. 2003.

H. B. Liu, B. C. Hsieh, P. T. P. Ho, L. Lin, and R. Yan. A New Galaxy Group Finding Algorithm: Probability Friends-of-Friends. ApJ, 681:1046–1057, July 2008.

E. J. Lloyd-Davies, A. K. Romer, M. Hosmer, N. Mehrtens, M. Davidson, K. Sabirli, R. G. Mann, and Hilton, M. et al. The XMM Cluster Survey: X-ray analysis methodology. *ArXiv e-prints*, October 2010.

M. S. Longair and M. Seldner. The clustering of galaxies about extragalactic radio sources. MNRAS, 189:433–453, November 1979.

P. A. A. Lopes, R. R. de Carvalho, R. R. Gal, S. G. Djorgovski, S. C. Odewahn, A. A. Mahabal, and R. J. Brunner. The Northern Sky Optical Cluster Survey. IV. An Intermediate-Redshift Galaxy Cluster Catalog and the Comparison of Two Detection Algorithms. AJ, 128:1017–1045, September 2004.

J. R. Lucey. An assessment of the completeness and correctness of the Abell catalogue. MNRAS, 204:33–43, July 1983.

S. L. Lumsden, R. C. Nichol, C. A. Collins, and L. Guzzo. The Edinburgh-Durham Southern Galaxy Catalogue. IV - The Cluster Catalogue. MNRAS, 258:1–22, September 1992.

S. J. Maddox, G. Efstathiou, W. J. Sutherland, and J. Loveday. The APM galaxy survey. I - APM measurements and star-galaxy separation. MNRAS, 243:692–712, April 1990.

C. Maraston, G. Strömbäck, D. Thomas, D. A. Wake, and R. C. Nichol. Modelling the colour evolution of luminous red galaxies - improvements with empirical stellar spectra. MNRAS, 394:L107–L111, March 2009.

M. Markevitch. Chandra Observation of the Most Interesting Cluster in the Universe. In A. Wilson, editor, *The X-ray Universe 2005*, volume 604 of *ESA Special Publication*, pages 723–+, January 2006.

B. J. Maughan, C. Jones, W. Forman, and L. Van Speybroeck. Images, Structural Properties, and Metal Abundances of Galaxy Clusters Observed with Chandra ACIS-I at $0.1<z<1.3$. ApJS, 174:117–135, January 2008.

I. G. McCarthy, C. S. Frenk, A. S. Font, C. G. Lacey, R. G. Bower, N. L. Mitchell, M. L. Balogh, and T. Theuns. Ram pressure stripping the hot gaseous haloes of galaxies in groups and clusters. MNRAS, 383:593–605, January 2008.

I. G. McCarthy, J. Schaye, R. G. Bower, T. J. Ponman, C. M. Booth, C. Dalla Vecchia, and V. Springel. Gas expulsion by quasar-driven winds as a solution to the overcooling problem in galaxy groups and clusters. MNRAS, page 35, January 2011.

N. Mehrtens, A. K. Romer, E. J. Lloyd-Davies, M. Hilton, C. J. Miller, and Stanford, S. A. et al. The XMM Cluster Survey: Optical analysis methodology and the first data release. *ArXiv e-prints*, June 2011.

A. J. Metevier, A. K. Romer, and M. P. Ulmer. The Butcher-Oemler Effect at Moderate Redshift. AJ, 119:1090–1099, March 2000.

C. J. Miller, R. C. Nichol, D. Reichart, R. H. Wechsler, A. E. Evrard, J. Annis, T. A. McKay, and Bahcall, N. A. et al. The C4 Clustering Algorithm: Clusters of Galaxies in the Sloan Digital Sky Survey. AJ, 130:968–1001, September 2005.

S. Miyazaki, T. Hamana, K. Shimasaku, H. Furusawa, M. Doi, M. Hamabe, K. Imi, and Kimura, M. et al. Searching for Dark Matter Halos in the Suprime-Cam 2 Square Degree Field. ApJ, 580:L97–L100, December 2002.

D. G. Monet, S. E. Levine, B. Canzian, H. D. Ables, A. R. Bird, C. C. Dahn, H. H. Guetter, and H. C. et al. Harris. The USNO-B Catalog. AJ, 125:984–993, February 2003.

R. E. M. Moore and I. O. Angell. Voronoi Polygons and Polyhedra. *Journal of Computational Physics*, 105:301–305, April 1993.

J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. ApJ, 490:493, December 1997.

A. B. Newman, R. S. Ellis, T. Treu, and K. Bundy. Keck Spectroscopy of z $>$ 1 Field Spheroidals: Dynamical Constraints on the Growth Rate of Red "Nuggets". ApJ, 717: L103–L107, July 2010.

R. C. Nichol, R. K. Sheth, Y. Suto, A. J. Gray, I. Kayo, R. H. Wechsler, F. Marin, and Kulkarni, G. et al. The effect of large-scale structure on the SDSS galaxy three-point correlation function. MNRAS, 368:1507–1514, June 2006.

M. R. Nolta, J. Dunkley, R. S. Hill, G. Hinshaw, E. Komatsu, D. Larson, L. Page, and Spergel, D. N et al. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Angular Power Spectra. ApJS, 180:296–305, February 2009.

P. Norberg, S. Cole, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, and Cannon, R. et al. The 2dF Galaxy Redshift Survey: the bJ-band galaxy luminosity function and survey selection function. MNRAS, 336:907–931, November 2002.

D. Novikov, S. Colombi, and O. Doré. Skeleton as a probe of the cosmic web: the two-dimensional case. MNRAS, 366:1201–1216, March 2006.

C. Park, Y.-Y. Choi, M. S. Vogeley, J. R. I. Gott, J. Kim, C. Hikage, T. Matsubara, and Park, M.-G. et al. Topology Analysis of the Sloan Digital Sky Survey. I. Scale and Luminosity Dependence. ApJ, 633:11–22, November 2005.

S. V. Penton, J. T. Stocke, and J. M. Shull. The Local Lyα Forest. IV. Space Telescope Imaging Spectrograph G140M Spectra and Results on the Distribution and Baryon Content of H I Absorbers. ApJS, 152:29–62, May 2004.

W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, and Gunn, J. E. et al. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. MNRAS, 401:2148–2168, February 2010.

S. Perlmutter, G. Aldering, M. della Valle, S. Deustua, R. S. Ellis, and Fabbro, S. et al. Discovery of a supernova explosion at half the age of the universe. Nature, 391:51–+, January 1998.

K. A. Pimbblet and M. J. Drinkwater. Intercluster Filaments of Galaxies Programme: pilot study survey and results. MNRAS, 347:137–143, January 2004.

K. A. Pimbblet, M. J. Drinkwater, and M. C. Hawkrigg. Intercluster filaments of galaxies programme: abundance and distribution of filaments in the 2dFGRS catalogue. MNRAS, 354:L61–L65, November 2004.

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, and Aumont, J. et al. Planck Early Results: The all-sky Early Sunyaev-Zeldovich cluster sample. *ArXiv e-prints*, January 2011.

M. Plionis, J. D. Barrow, and C. S. Frenk. Projected and intrinsic shapes of galaxy clusters. MNRAS, 249:662–677, April 1991.

M. Postman, L. M. Lubin, J. E. Gunn, J. B. Oke, J. G. Hoessel, D. P. Schneider, and J. A. Christensen. The Palomar Distant Clusters Survey. I. The Cluster Catalog. AJ, 111:615, February 1996.

E. A. Praton, A. L. Melott, and M. Q. McKee. The Bull's-Eye Effect: Are Galaxy Walls Observationally Enhanced? ApJ, 479:L15+, April 1997.

M. Ramella, W. Boschin, D. Fadda, and M. Nonino. Finding galaxy clusters using Voronoi tessellations. A&A, 368:776–786, March 2001.

A. C. S. Readhead, B. S. Mason, C. R. Contaldi, T. J. Pearson, J. R. Bond, S. T. Myers, S. Padin, and Sievers, J. L., et al. Extended Mosaic Observations with the Cosmic Background Imager. ApJ, 609:498–512, July 2004.

C. L. Reichardt, P. A. R. Ade, J. J. Bock, J. R. Bond, J. A. Brevik, C. R. Contaldi, M. D. Daub, and Dempsey, J.T. et al. High-Resolution CMB Power Spectrum from the Complete ACBAR Data Set. ApJ, 694:1200–1219, April 2009.

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, and Garnavich, P. M. et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. AJ, 116:1009–1038, September 1998.

A. G. Riess, L. Macri, S. Casertano, M. Sosey, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, and Jha, S. W. et al. A Redetermination of the Hubble Constant with the Hubble Space Telescope from a Differential Distance Ladder. ApJ, 699:539–563, July 2009.

K. Rines and M. J. Geller. Spectroscopic Determination of the Luminosity Function in the Galaxy Clusters A2199 and Virgo. AJ, 135:1837–1848, May 2008.

A. J. Romanowsky, N. G. Douglas, M. Arnaboldi, K. Kuijken, M. R. Merrifield, N. R. Napolitano, M. Capaccioli, and K. C. Freeman. A Dearth of Dark Matter in Ordinary Elliptical Galaxies. *Science*, 301:1696–1698, September 2003.

A. K. Romer, P. T. P. Viana, A. R. Liddle, and R. G. Mann. A Serendipitous Galaxy Cluster Survey with XMM: Expected Catalog Properties and Scientific Applications. ApJ, 547: 594–608, February 2001.

E. Rozo, E. S. Rykoff, B. P. Koester, T. McKay, J. Hao, A. Evrard, R. H. Wechsler, and Hansen, S. et al. An Improved Cluster Richness Estimator. *astro-ph/0809.2797*, September 2008.

V. C. Rubin. Differential rotation of the inner metagalaxy. AJ, 56:47–+, 1951.

W. E. Schaap and R. van de Weygaert. Continuous fields and discrete samples: reconstruction through Delaunay tessellations. A&A, 363:L29–L32, November 2000.

C. Scharf, M. Donahue, G. M. Voit, P. Rosati, and M. Postman. Evidence for X-Ray Emission from a Large-Scale Filament of Galaxies? ApJ, 528:L73–L76, January 2000.

P. Schechter. An analytic expression for the luminosity function for galaxies. ApJ, 203: 297–306, January 1976.

D. J. Schlegel, D. P. Finkbeiner, and M. Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. ApJ, 500:525, June 1998.

Frederic Paik Schoenberg, Christopher Barr, and Jungju Seo. The distribution of voronoi cells generated by southern california earthquake epicenters. *Environmetrics*, 20(2): 159–171, 2009. ISSN 1099-095X. URL `http://dx.doi.org/10.1002/env.917`.

M. Seldner, B. Siebers, E. J. Groth, and P. J. E. Peebles. New reduction of the Lick catalog of galaxies. AJ, 82:249–256, April 1977.

C. D. Shane and C. A. Wirtanen. The distribution of galaxies. *Publ. Lick Obs*, 22, 1967.

S. A. Shectman. Clusters of galaxies from the Shane-Wirtanen counts. ApJS, 57:77–90, January 1985.

Robin Sibson. The dirichlet tessellation as an aid in data analysis. *Scandinavian Journal of Statistics*, 7(1):pp. 14–20, 1980. ISSN 03036898. URL `http://www.jstor.org/stable/4615765`.

M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, and Capps, R. et al. The Two Micron All Sky Survey (2MASS). AJ, 131: 1163–1183, February 2006.

R. J. Smith, J. R. Lucey, M. J. Hudson, S. P. Allanson, T. J. Bridges, A. E. Hornschemeier, R. O. Marzke, and N. A. Miller. A spectroscopic survey of dwarf galaxies in the Coma cluster: stellar populations, environment and downsizing. MNRAS, 392:1265–1294, February 2009.

R. M. Soneira and P. J. E. Peebles. A computer model universe - Simulation of the nature of the galaxy distribution in the Lick catalog. AJ, 83:845–849, July 1978.

T. Sousbie, S. Colombi, and C. Pichon. The fully connected N-dimensional skeleton: probing the evolution of the cosmic web. MNRAS, 393:457–477, February 2009.

T. Sousbie, C. Pichon, S. Colombi, D. Novikov, and D. Pogosyan. The 3D skeleton: tracing the filamentary structure of the Universe. MNRAS, 383:1655–1670, February 2008a.

T. Sousbie, C. Pichon, H. Courtois, S. Colombi, and D. Novikov. The Three-dimensional Skeleton of the SDSS. ApJ, 672:L1–L4, January 2008b.

D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, and Dunkley, J. et al. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. ApJS, 170:377–408, June 2007.

V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, and Frenk, C. S. et al. The Aquarius Project: the subhaloes of galactic haloes. MNRAS, 391:1685–1711, December 2008.

V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, and Thacker, R. et al. Simulations of the formation, evolution and clustering of galaxies and quasars. Nature, 435:629–636, June 2005.

Z. Staniszewski, P. A. R. Ade, K. A. Aird, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, and Cho, H.-M. et al. Galaxy Clusters Discovered with a Sunyaev-Zel'dovich Effect Survey. ApJ, 701:32–41, August 2009.

R. S. Stoica, V. J. Martínez, J. Mateu, and E. Saar. Detection of cosmic filaments using the Candy model. A&A, 434:423–432, May 2005.

R. S. Stoica, V. J. Martínez, and E. Saar. Filaments in observed and mock galaxy catalogues. A&A, 510:A38+, February 2010.

J. P. Stott. *The Evolution of Galaxies in Massive Clusters*. PhD thesis, Durham University, 2007.

J. P. Stott, K. A. Pimbblet, A. C. Edge, G. P. Smith, and J. L. Wardlow. The evolution of the red sequence slope in massive galaxy clusters. MNRAS, 394:2098–2108, April 2009.

M. F. Struble and H. J. Rood. A compilation of redshifts and velocity dispersions for Abell clusters. ApJS, 63:543–553, March 1987.

W. Sutherland. The 3-D distribution of Abell clusters. MNRAS, 234:159–172, September 1988.

E. Tago, J. Einasto, E. Saar, E. Tempel, M. Einasto, J. Vennik, and V. Müller. Groups of galaxies in the SDSS Data Release 5. A group-finder and a catalogue. A&A, 479: 927–937, March 2008.

M. Tanemura. Statistical Distributions of Poisson Voronoi Cells in Two and Three Dimensions. *FORMA*, 18:221–247, November 2003.

M. B. Taylor. TOPCAT & STIL: Starlink Table/VOTable Processing Software. In P. Shopbell, M. Britton, & R. Ebert, editor, *Astronomical Data Analysis Software and Systems XIV*, volume 347 of *Astronomical Society of the Pacific Conference Series*, pages 29–+, December 2005.

M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, and Weinberg, D. H. et al. The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey. ApJ, 606:702–740, May 2004.

C. van Breukelen and L. Clewley. A reliable cluster detection technique using photometric redshifts: introducing the 2TecX algorithm. MNRAS, 395:1845–1856, June 2009.

L. Verde. Non-Gaussianity from Large-Scale Structure Surveys. *Advances in Astronomy*, 2010.

W. Voges, B. Aschenbach, T. Boller, H. Bräuninger, U. Briel, W. Burkert, K. Dennerl, and Englhauser, J. et al. The ROSAT all-sky survey bright source catalogue. A&A, 349: 389–405, September 1999.

Georges Voronoi. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1908(134):198–287, January 1908. URL `http://dx.doi.org/10.1515/crll.1908.134.198`.

S. Weinberg. The cosmological constant problem. *Reviews of Modern Physics*, 61:1–23, January 1989.

N. Werner, A. Finoguenov, J. S. Kaastra, A. Simionescu, J. P. Dietrich, J. Vink, and H. Böhringer. Detection of hot gas in the filament connecting the clusters of galaxies Abell 222 and Abell 223. A&A, 482:L29–L33, May 2008.

D. Wittman, I. P. Dell'Antonio, J. P. Hughes, V. E. Margoniner, J. A. Tyson, J. G. Cohen, and D. Norman. First Results on Shear-selected Clusters from the Deep Lens Survey: Optical Imaging, Spectroscopy, and X-Ray Follow-up. ApJ, 643:128–143, May 2006.

C. Wolf, K. Meisenheimer, M. Kleinheinrich, A. Borch, S. Dye, M. Gray, L. Wisotzki, and Bell, E. F. et al. A catalogue of the Chandra Deep Field South with multi-colour classification and photometric redshifts from COMBO-17. A&A, 421:913–936, July 2004.

G. Worthey, S. C. Trager, and S. M. Faber. The Galaxian Age-Metallicity Relation. In A. Buzzoni, A. Renzini, & A. Serrano, editor, *Fresh Views of Elliptical Galaxies*, volume 86 of *Astronomical Society of the Pacific Conference Series*, page 203, 1995.

Y. C. Yaryura, C. M. Baugh, and R. E. Angulo. Are the 2dFGRS superstructures a problem for hierarchical models? *arXiv:astro-ph/0007281*, March 2010.

H. K. C. Yee and E. Ellingson. Correlations of Richness and Global Properties in Galaxy Clusters. ApJ, 585:215–226, March 2003.

H. K. C. Yee and O. López-Cruz. A Quantitative Measure of the Richness of Galaxy Clusters. AJ, 117:1985–1994, May 1999.

D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, and Barkhouser, R. et al. The Sloan Digital Sky Survey: Technical Summary. AJ, 120:1579–1587, September 2000.

H. Yu, P. Tozzi, S. Borgani, P. Rosati, and Z.-H. Zhu. Measuring redshifts using X-ray spectroscopy of galaxy clusters: results from Chandra data and future prospects. A&A, 529:A65+, May 2011.

I. Zehavi, M. R. Blanton, J. A. Frieman, D. H. Weinberg, H. J. Mo, M. A. Strauss, S. F. Anderson, and Annis, J. et al. Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data. ApJ, 571:172–190, May 2002.

Y. Zhang, X. Yang, A. Faltenbacher, V. Springel, W. Lin, and H. Wang. The Spin and Orientation of Dark Matter Halos Within Cosmic Filaments. ApJ, 706:747–761, November 2009.

F. Zwicky. On the Masses of Nebulae and of Clusters of Nebulae. ApJ, 86:217, October 1937.

F. Zwicky, E. Herzog, and P. Wild. Catalogue of galaxies and of clusters of galaxies, Vol. I. *Catalogue of galaxies and of clusters of galaxies, Vol. I, Pasadena: California Institute of Technology (CIT), —c1961*, 1961.